ISSN 2587-814X (print), ISSN 2587-8158 (online)
Russian version: ISSN 1998-0663 (print), ISSN 2587-8166 (online)

Vol. 17 No. 4 – 2023

BUSINESS INFORMATICS

HSE Scientific Journal

Publisher:
HSE University

The journal is published quarterly

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation

Address:
26-28, Shabolovka Street,
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 *28509
http://bijournal.hse.ru
E-mail: bijournal@hse.ru

Circulation:
English version – 100 copies,
Russian version – 100 copies,
online versions in English and Russian – open access

Printed in HSE Printing House
44, build.2, Izmaylovskoye Shosse,
Moscow, Russia

CONTENTS

M.B. Laskin, O.V. Rusakov
Prediction of distributions of unit prices for real estate properties on the basis of the characteristics of PSI-processes 7

A.V. Kudrov
The impact of economic complexity and industry specialization on the gross regional product of Russian regions 25

R.S. Rogulin
Mathematical model of the formation of supply chains of raw materials from a commodity exchange under conditions of uncertainty 41

I.V. Solovyov, V.A. Semenikhin, S.P. Kushch
The influence of the breadth of the tech stack on the result of the digital product: A view through the theory of absorption capacity 57

Yu.F. Telnov, V.A. Kazakov, A.A. Bryzgalov, I.G. Fiodorov
Methods and models for substantiating application scenarios for the digitalization of manufacturing and business processes of network enterprises 73

M.K. Pandey, A.K.Pathak, I.G. Sergeeva
A bibliometric review of scientific research on the significance of information technology relating to sustainable development reporting practice 94

ABOUT THE JOURNAL

Business Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by HSE University, Moscow, Russian Federation. The journal is administered by HSE Graduate School of Business. The journal is issued quarterly, in English and Russian.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the following areas: modeling of social and economic systems, digital transformation of business, innovation management, information systems and technologies in business, data analysis and business intelligence systems, mathematical methods and algorithms of business informatics, business processes modeling and analysis, decision support in management.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation.

The journal is included into Scopus, Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index on the Web of Science platform (RSCI), EBSCO.

The journal is distributed both in printed and electronic forms.

EDITORIAL BOARD

ABOUT HSE UNIVERSITY

Consistently ranked as one of Russia's top universities, HSE University is a leader in Russian education and one of the preeminent economics and social sciences universities in Eastern Europe and Eurasia.

Having rapidly grown into a well-renowned research university over two decades, HSE sets itself apart with its international presence and cooperation.

Our faculty, researchers, and students represent over 50 countries, and are dedicated to maintaining the highest academic standards. Our newly adopted structural reforms support both HSE's drive to internationalize and the groundbreaking research of our faculty, researchers, and students.

Now a dynamic university with four campuses, HSE is a leader in combining Russian educational traditions with the best international teaching and research practices. HSE offers outstanding educational programs from secondary school to doctoral studies, with top departments and research centers in a number of international fields.

Since 2013, HSE has been a member of the 5-100 Russian Academic Excellence Project, a highly selective government program aimed at boosting the international competitiveness of Russian universities.

ABOUT HSE GRADUATE SCHOOL OF BUSINESS

HSE Graduate School of Business was created on September 1, 2020. The School will become a priority partner for leading Russian companies in the development of their personnel and management technologies.

The world-leading model of a ‘university business school’ has been chosen for the Graduate School of Business. This foresees an integrated portfolio of programmes, ranging from Bachelor’s to EMBA programmes, communities of experts and a vast network of research centres and laboratories for advanced management studies. Furthermore, HSE University’s integrative approach will allow the Graduate School of Business to develop as an interdisciplinary institution. The advancement of the Graduate School of Business through synergies with other faculties and institutes will serve as a key source of its competitive advantage. Moreover, the evolution and development of the Business School’s faculty involves the active engagement of three professional tracks at our University: research, practice-oriented and methodological.

What sets the Graduate School of Business apart is its focus on educating and developing globally competitive and socially responsible business leaders for Russia’s emerging digital economy.

The School’s educational model will focus on a project approach and other dynamic methods for skills training, integration of online and other digital technologies, as well as systematic internationalization of educational processes.

At its start, the Graduate School of Business will offer 22 Bachelor programmes (three of which will be fully taught in English) and over 200 retraining and continuing professional development programmes, serving over 9,000 students. In future, the integrated portfolio of academic and professional programmes will continue to expand with a particular emphasis on graduate programmes, which is in line with the principles guiding top business schools around the world. In addition, the School’s top quality and all-encompassing Bachelor degrees will continue to make valuable contributions to the achievement of the Business School’s goals and the development t of its business model.

The School’s plans include the establishment of a National Resource Center, which will offer case studies based on the experience of Russian companies. In addition, the Business School will assist in the provision of up-to-date management training at other Russian universities. Furthermore, the Graduate School of Business will become one of the leaders in promoting Russian education.

The Graduate School of Business’s unique ecosystem will be created through partnerships with leading global business schools, as well as in-depth cooperation with firms and companies during the entire life cycle of the school’s programmes. The success criteria for the Business School include professional recognition thanks to the stellar careers of its graduates, its international programmes and institutional accreditations, as well as its presence on global business school rankings.

DOI: 10.17323/2587-814X.2023.4.7.24

Prediction of distributions of unit prices for real estate properties on the basis of the characteristics of PSI-processes*

Mikhail B. Laskin [a]
E-mail: laskinmb@yahoo.com

Oleg V. Rusakov [b]
E-mail: ovirusakov@yahoo.co.uk

[a] Saint Petersburg Federal Research Center, Russian Academy of Sciences
Address: 39, 14th Line, Vasilyevsky Island, St. Petersburg 199178, Russia

[b] Saint Petersburg State University
Address: 7/9, Universitetskaya Nab., St. Petersburg 199034, Russia

Abstract

Real estate market price forecasting is always in the focus of interests of scientists-economists, market analysts, market participants (sellers and buyers), marketing services of building complex enterprises, analysts working for banks and insurance companies and investors. Under present day conditions, the price behavior of properties on real estate markets takes especially important meaning subject to the influence of such factors as changes in the structure of household incomes, changes in mortgage rates and their availability, dynamic changes in the macroeconomic and other external socio-economic and

* The article is published with the support of the HSE University Partnership Programme

political type factors. However, unlike the financial and securities markets, the real estate market is always characterized by a delayed reaction to external perturbations, often up to half a year, which allows us to hope for an appropriate construction of forecasts, at least in time for the delayed reaction. Traditional autoregressive forecasting methods are characterized by rapidly increasing forecast variance, because they assume a factor of stochastic volatility. This paper proposes a model and method of forecast construction based on stochastic processes of the "Poisson random index" having a short time for reaching a stationary stable variance. The model is based on the "principle of replacements" of current prices with new ones. We analyze in detail an example of the application of the "principle of replacements" for construction of price forecasts on secondary residential real estate in St. Petersburg which is based on data of four-year observations of offer prices.

Keywords: real estate price forecast, log-normal price distribution, pseudo-Poisson process, Poisson random index process, Ornstein–Uhlenbeck process

Citation: Laskin M.B., Rusakov O.V. (2023) Prediction of distributions of unit prices for real estate properties on the basis of the characteristics of PSI-processes. *Business Informatics*, vol. 17, no. 4, pp. 7–24.
DOI: 10.17323/2587-814X.2023.4.7.24

Introduction

The real estate market is an important sector of the economy of any country. Capital construction projects entering the turnover of the commodity real estate market create chains of economic relations in the construction industry, in the sectors of construction materials production, and in the extractive industry. The sphere of turnover of primary and secondary real estate properties is to a large extent the area of business interests of the banking and insurance sectors of the economy. The aggregate of real estate properties owned and in commercial turnover is the taxable base for property tax and, consequently, a basis for replenishment of the country's budget. For these reasons, forecasting prices in the real estate market has always been and remains an urgent task for all market participants. Materials devoted to this subject regularly appear in periodicals. The tasks related to the construction of forecasts are constantly the subject of research by scientists and researchers. Among the works of a general methodological nature, it is worth mentioning the well-known translated book [1], a series of works by the authors of publications [2–5]. As examples of relatively recent publications of domestic researchers, we can name works [6–16] and foreign works – [17–22]. Real estate markets, as rightly noted in [5], have characteristic regional features. The works of domestic authors [3, 6, 10, 11, 13–16, 23, 24] and foreign authors – [25–28] are devoted to price forecasting in regional domestic markets. In most cases, traditional forecasting methods are considered: fundamental and technical analysis [11], factor models [14], regression models, autoregressive and moving average models [2–5, 15, 16, 18, 27]. At the same time, "standard econometric methods are unsuitable for forecasting real estate market trends in modern conditions" and "methods developed in countries with developed market economies are unsuitable for forecasting in countries with transition economies" [5]. Recently, there are works in which machine learning methods are applied for the purposes of real estate price forecasting [24], including neural network modeling methods [8, 29]. The main disadvantage of traditional methods based on autoregression is a rapid increase in the forecast variance which makes the forecast result uninformative after 2–3 steps. This is due to the introduction of a random scale of volatility, which is reflected in the fact that a mixture of distributions arises – and the variance of the mixture is always greater than or equal to the mixture of variances. At the same time, numerous empirical observations suggest that the variance of unit prices of relatively similar real estate properties changes little over time, even in the presence of strong upward or downward trends. The exception may be relatively short periods of exposure to strong external perturbing factors leading to a noticeable change in the trend. In this regard, a forecast with a relatively stable variance would be preferable.

In this paper, we propose a mathematical model of the price change process in the real estate market under simple and seemingly natural assumptions. Such assumptions are the following statements:

1) the announced selling price for some property remains unchanged during some (random) time interval;
2) at any point in time, the property may be withdrawn from sale or sold;
3) at any moment of time, a new property may appear in the listing of properties for sale and replace the property withdrawn from sale or sold;
3) the prices of the properties in the listing at each fixed point in time are independent, or at least conditionally independent subject to some external factor;
4) the total set of properties for sale at a fixed point in time forms an observable sample of unit prices, which is subject to study and statistical processing.

The theoretical basis of the model proposed below is based on three provisions:

- the principle of log-normal distribution of unit prices of relatively homogeneous real estate properties;
- characteristics of the distribution and covariance function of the Poisson random index process (hereinafter referred to as PSI-process);
- the central limit theorem for PSI-processes – convergence of their normalized sums to the stochastic Ornstein–Uhlenbeck process (stationary, Gaussian, Markov process).

In [30, 31] the justification of the convergence of unit prices formed by successive comparisons to a log-normal distribution is given. Apparently, the first work in which noted the adherence of unit rental rates to a log-normal distribution was the work of British statisticians [32]. This adherence is also noted in the work of Japanese scientists [26]. In [33] the characteristics of distributions and correlation functions of PSI-processes are investigated, and the convergence of normalized sums of independent PSI-processes to Ornstein–Uhlenbeck type processes is proved there. The theoretical foundations of the model proposed in this paper are presented in [30, 31, 33, 34].

1. Definition and basic characteristics of the PSI-process

Let $\xi_0, \xi_1, \xi_2, \ldots$ be a random sequence, which we will call the forming or slave sequence; $\Pi(t) = \Pi_\lambda(t)$, $t \geqslant 0$ is a Poisson process independent of it with intensity $\lambda > 0$, which we will call the master. We define a Poisson subordinator1 for the sequence $\xi_0, \xi_1, \xi_2, \ldots$ as follows $\psi(t) = \psi_\lambda(t) := \xi_{\Pi(t)}$. The resulting process $\psi(t)$ with continuous time $t \geqslant 0$ – we will call the Poisson Stochastic Index process or PSI-process. Note that PSI-processes are a natural generalization of the pseudo-Poisson processes introduced and discussed in detail in Chapter X of the second volume of Feller's classic work [35].

The PSI-process represents successive replacements of the members of the forming sequence, occurring at the moments of jumps of the Poisson process. Time intervals $\{\tau_{j+1}\}, j = 0, 1, 2, \ldots$ between consecutive jumps of the leading Poisson process are called spacings. It is known that spacings are independent identically distributed random variables with a common exponential (or, what is the same, exponential) distribution having intensity $\lambda > 0$. At time zero, ξ_0, which "holds" its value during the first spacing, is played out, at the time of the first jump of the Poisson process it is replaced by ξ_1, and so on.... During the j-th spacing, the played random variable $\{\xi_{j-1}\}$ does not change its value until (inclusive

1 Subordination in theories of stochastic processes is called the generally accepted random replacement of time.

of) the moment of the j-th jump of the Poisson process θ_j. At the moment of the spacing change τ_{j+1} to τ_{j+2}, the value of the random variable ξ_j is replaced by ξ_{j+1} (*Fig. 1*).

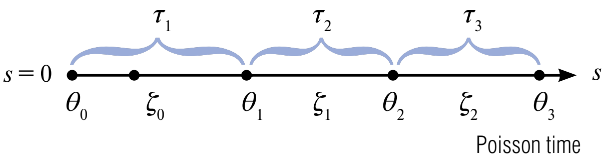


Fig. 1. Schematic representation of the random spacing lengths $\tau_0, \tau_1, \tau_2, \ldots$, the moments of the Poisson process jumps $\theta_0, \theta_1, \theta_2, \ldots$ and the substituted random variables $\xi_0, \xi_1, \xi_2, \ldots$.

Note that the PSI-process has the following representation as a weighted sum of elements of the random sequence $\xi_0, \xi_1, \xi_2, \ldots$, where the weights are Poisson indicators

$$\psi_\lambda(t) = \sum_{j=0}^{\infty} \xi_j \cdot \mathbf{1}\{\Pi(t) = j\},$$

hereafter **1** is the indicator function.

If $\xi_0, \xi_1, \xi_2, \ldots$ represent a stationary sequence, then the PSI-process is stationary. In particular, this will be fulfilled when $\xi_0, \xi_1, \xi_2, \ldots$ are independent identically distributed quantities. In the latter case, if $E\xi_0 = a$, then the conditional mathematical expectation of the PSI-process is

$$E\left(\psi_\lambda\left(t+s\right) \mid \psi_\lambda(s) = z\right) = ze^{-\lambda \cdot s} + a(1 - e^{-\lambda \cdot s})$$

for any arbitrary but fixed z, for any non-negative t, s. And, in particular, when $a = 0$

$$E\left(\psi_\lambda\left(t+s\right) \mid \psi_\lambda(s) = z\right) = ze^{-\lambda \cdot s}.$$

It is important to note that if the sequence $\xi_0, \xi_1, \xi_2, \ldots$ is Markovian, then the corresponding PSI-process also has the Markov characteristic[2].

2. The covariance function of the process $\psi(t)$

In [33] the following result is obtained. Let $\xi_0, \xi_1, \xi_2, \ldots$ are independent identically distributed random variables, $E(\xi_0) = 0$, $D(\xi_0) = 1$, then the covariance function of the process $\psi(t)$, $t \geqslant 0$ decreases exponentially and has the form

$$cov(\psi(t), \psi(0)) = \exp(-\lambda t).$$

Consider independent copies of a single PSI-process $\psi_\lambda(t)$: $\psi_1(t), \psi_2(t), \ldots$ It follows from the exponential form of the covariance function that, by virtue of the central limit theorem (CLT) for vectors, a random process composed of normalized sums of PSI-processes of the form

$$Z_N(t) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N} \psi_{(j)}(t),$$

converges in the sense of weak convergence of finite-dimensional distributions[3], when $N \to \infty$, to the Ornstein–Uhlenbeck process, a stationary, Gaussian, Markov process, and given in a "standardized" form. The latter means that at arbitrary moments of time $(t_1, \ldots, t_d)$, the vector $Z(t_1), \ldots, Z(t_d)$ has a joint normal distribution with zero mean and with covariance $\exp(-\lambda|t_i - t_j|)$, where t_i, t_j are running all elements of the set$(t_1, \ldots, t_d)$. Thus, the distribution of the "standardized" Ornstein–Uhlenbeck process is characterized by Gaussianity (joint normality of finite-dimensional distributions), zero mean (zero "theoretical" trend), and covariance of the form $\exp(-\lambda t)$. The coefficient $\lambda > 0$ is called the "speed" of the Ornstein–Uhlenbeck process, and $1/\lambda$ is called the "viscosity". Here we see that the rate of the limiting Ornstein–Uhlenbeck process is exactly the intensity of the "leading" Poisson process. The conditional expectation of the Ornstein–Uhlenbeck process coincides with the conditional expectation of the PSI-process

$$E\left(U(t+s) \mid U(t) = z\right) = ze^{-\lambda \cdot s},$$

[2] The Markov characteristic is when the future, with a fixed past and present, does not depend on the past (that is, it depends on the past only through the present).

[3] Moreover, there is convergence of $Z_N(t)$ in the Skorokhod functional space, see [34].

and the conditional variance is independent of the z condition and is equal to

$$D\big(U(t+s)\,|\,U(t)=z\big)=1-e^{-2\lambda\cdot s}.$$

Moreover, the conditional distribution of the process U given z is normal at any non-negative s. From the standard Ornstein–Uhlenbeck process by shifting a and scaling $b > 0$, we obtain an Ornstein–Uhlenbeck process OU which has a stationary distribution normal with parameters a and b^2, covariance $b^2\exp(-\lambda t)$. For such an Ornstein–Uhlenbeck process, of course, stationarity and the Markov property are preserved, and the conditional expectation and conditional variance are respectively equal to

$$E\big(OU(t+s)\,|\,OU(s)=y\big)=a+(y-a)e^{-\lambda\cdot s},$$
$$D\big(OU(t+s)\,|\,OU(t)=y\big)=b^2(1-e^{-2\lambda\cdot s}). \qquad (1)$$

From where, in particular, we can see that the conditional variance of OU is independent of the condition.

Applied to real estate prices, we interpret the model under consideration as follows. We represent the price listing as a table, where each row is a time slice of current offer prices (N is the volume of the slice, changes from slice to slice), and each column is the price of the property in dynamics, possibly with correction for trend (n is the number of slices). We observe price slices with some periodicity determined by the next issue of the price log (usually once a week). By the next issue of the log, each object may “go away” (or dramatically “go away” its price), or it may stay with the same price. A property, or its price, can be replaced by a newly arrived property (price). Moreover, the remaining prices from the previous issue of the journal are significantly more than the newly arrived ones in the subsequent issue. Each PSI-process is represented by a column in a table. The values are observed in time slices, according to the dates of the next issue of the journal. All PSI-processes are assumed to be independent and identically distributed (i.e., they are independent copies of a single PSI-process – in our context, price, or the logarithm of price). The distribution of each PSI-process is understood as the distribution of a piecewise constant function with continuous time, continuous on the right, having finite limits on the left (such functions are called Right Continuous Left Limits, RCLL). The Poisson process acts as a point process determining the moments of replacements due to its characteristic of “no aftereffect”: no matter how much time has passed since the previous jump, the next jump will occur after an exponential time.

In our approach, we make the following ***approximation of the logarithms of prices*** in the “table”. Let $V(t)$ be the price of 1 square meter of the selected property type at time t. In each slice at time t we perform logarithmization and observe independent realizations of the process $\ln(V(t))$ with added trend and scale factor $\sigma > 0$. Thus, at each time slice t we have (our) simple sample of size equal to the number of prices in the given time slice.

The method of constructing forecasts of price distributions and their numerical characteristics, proposed in this paper, is based on specific aspects of PSI-processes and limits of their normalized sums: Ornstein–Uhlenbeck processes. The main general properties here (for both the PSI-process and the Ornstein–Uhlenbeck process) are Markovianness, the kind of conditional mathematical expectations. We also use the type of conditional variance and the Gaussian property of the conditional distribution of the Ornstein–Uhlenbeck process.

Constant monitoring of prices on the real estate market, with a predetermined periodicity, allows us to use the characteristics of the Ornstein–Uhlenbeck processes to forecast future price distributions and their numerical characteristics, such as mean, median, modal (market value) values, standard deviation, corridors of acceptable values, and so on. The peculiarity of the real estate market is that price changes on the market can be observed, as a rule, not more than once a week, as most printed advertising publications and Internet resources are updated with the same frequency. At the same time, the real estate market has a slow reaction to external sources of disturbances of macroeconomic nature, as the search for an object, reaching agreements, registration of the transaction, entry into the ownership rights require considerable time. In this regard, it is quite reasonable to monitor prices with a periodicity of once a month, because during such a period of time price changes become noticeable.

3. Model

Let us consider the random process $V(t)$ – the dynamics of the price of 1 sq. m. of real estate in time and the associated process of the price logarithm $Y(t) = \ln(V(t))$. We take the average of each slice as the basic estimate of $Y(t)$ and denote it by the same letter (this also applies to $\hat{Y}(t)$, introduced below). Suppose the process $Y(t)$ has a linear trend $E(Y(t)) = \alpha \cdot t + \beta$ and a time-constant standard deviation $\sigma(t) \equiv \sigma > 0$. We will consider the centered and normalized process

$$\hat{Y}(t) = \frac{Y(t) - \alpha \cdot t - \beta}{\sigma}$$

as a PSI-process. Note that under the formulated assumptions $\hat{Y}(t)$ has a mathematical expectation identically equal to zero and a variance equal to one. Under the condition that the distribution (as a random process) of $\hat{Y}(t)$ coincides with the distribution of the PSI-process, or Ornstein–Uhlenbeck process, the covariance function is $cov\left(\hat{Y}(t)(s+t), \hat{Y}(t)\right) = \exp(-\lambda s)$ and does not depend on t. Since observations of prices in the real estate market are only available in discrete time, we speak of the values and observations of the process $V(t)$ at discrete points in time $V(j)$, $j = \overline{0, n-1}$. In [33, 34], weak convergence of the distribution of prices formed by successive comparisons to a log-normal distribution was proved. Thus, we have reasons to consider the sequence of price logarithms $Y(j) = \ln(V(j))$, $j = \overline{0, n-1}$ as a sequence of normally distributed random variables, whence it follows that all members of the corresponding sequence $\hat{Y}(t)$ are standardly normal. In this context, the physical meaning of the members of the sequence $(\xi_j), j = \overline{0, \infty}$ (from the definition of the PSI-process) can be the centered, standard deviation normalized logarithms of the price of 1 sq. m. of some property maintaining its value during the time interval (τ_i), $i = \overline{1, \infty}$, $(i = j + 1)$, i.e., a sequence of the form $\hat{Y}(j)$, $j = \overline{0, n-1}$. Under the assumption that the mean values of the price logarithms follow the Ornstein–Uhlenbeck process, for any pair of mean values of the logarithms of the random variables $V(t)$, $V(t+s)$, the hypothesis of a multivariate joint normal distribution for $\ln(V(t))$ can be considered. Note that quantiles (in particular, median) and mode can be considered as mean values here. If this hypothesis is confirmed for any predetermined $V(0) = v(0)$, forecast estimates can be obtained for the modal, median and mean values of the random variable $V(s)$ using the formulas of conditional mode, median and/or conditional mathematical expectation (see, for example, [36, 37]). Here, time 0 corresponds to the moment of the last observed price distribution, s is the time for which the forecast is given, its counting starts from the moment of the last observed distribution.

To verify the proposed interpretation, the following statements should be statistically proved and confirmed.

1) Study the distribution of spacings (τ_i), where $i = \overline{1, n}$ – number of periods of continuous price presence in the flow (the purpose is to obtain statistical confirmation of the hypothesis about their exponential distribution, to estimate the exponential distribution parameter).
2) Study price distributions $V(j)$, $j = \overline{0, n-1}$ in each slice (the purpose is to be convinced of the log-normal form of the price distribution);
3) Obtain confirmation of the independence of the copies of the random sequence $\hat{Y}(j)$, $j = \overline{0, n-1}$ for which purpose we study the behavior of the accumulated variances in each slice. We will check for uncorrelatedness, which together with Gaussianity will give independence.
4) Verify the characteristic of joint normality (Gaussianity) of the mean (for each slice) values of $\hat{Y}(t)$, which, together with the condition on the exponential form of the covariance function, confirms the Markov property of the random sequence composed of the mean (for each slice) of $\hat{Y}(j)$, $j = \overline{0, n-1}$. This follows from the fact that a stationary Gaussian random process with correlation decreasing exponentially is an Ornstein–Uhlenbeck process, and hence it has the Markov property. Moreover, the Gaussianity of the mean will confirm the Markov property of the median and mode as a function of $j = \overline{0, n-1}$.

Note that the Gaussianity of the mean for each slice is equal to the Gaussianity of the sums for each slice normalized by $\sqrt{N}$ for our range of values of N:

from 254 to 729. Also note that the Gaussianity of the averages may follow from the fact that the Ornstein–Uhlenbeck process in discrete time is a 1st order autoregression, and for uncorrelated noise it is sufficient to show conformity to the normal law.

4. Methodology of forecast construction[4]

1) Construct a series from the ***mean values*** of all time slices, identify the trend, subtract it from the series, obtain an estimate of the centered and standardized series $\hat{Y}(j)$, $j=\overline{0, n-1}$.
2) For the estimated series $\hat{Y}(j)$, $j=\overline{0, n-1}$, construct the partial autocovariance function (PACF), obtain confirmation of the hypothesis of conditional uncorrelation of $\hat{Y}(j)$ with $\hat{Y}(k)$, $|k-j|>0$.
3) For the estimated series $\hat{Y}(j)$, $j=\overline{0, n-1}$, construct the covariance function (autocovariance function, ACF), obtain confirmation of the hypothesis about the exponential form of the covariance function, estimate the exponent's degree parameter, and compare it with the exponential distribution parameter of the observed spacings.
4) Using formulas of the form (2)–(5) written below, construct a forecast.

Let us introduce the notations:

$E(\ln(V(0))) = \mu(0)$, $E(\ln(V(s))) = \mu(s)$,

$\sigma(\ln(V(0))) = \sigma(0)$, $\sigma(\ln(V(s))) = \sigma(s)$,

$\rho(\ln(V(0)), \ln(V(s))) = e^{-\lambda \cdot s}$.

Under the assumption that there is a linear trend of the form $\mu(s)=\alpha \cdot s+\beta$ for the mean logarithms of prices, the standard deviation is constant $\sigma(s) = \sigma(0) = \sigma$, we obtain for any predetermined value $V(0) = v(0)$ the forecast estimates:

$$Mode(V(s)|V(0) = v(0)) = \\ = \exp(\alpha \cdot s + \beta + e^{-\lambda \cdot s}(\ln(v(0)) - \mu(0) - \\ - \sigma^2(1 - e^{-2\lambda \cdot s})), \quad (2)$$

$$Median(V(s)|V(0) = v(0)) = \\ = \exp(\alpha \cdot s + \beta + e^{-\lambda \cdot s}(\ln(v(0)) - \mu(0))), \quad (3)$$

$$E(V(s)|V(0) = v(0)) = \\ = \exp(\alpha \cdot s + \beta + e^{-\lambda \cdot s}(\ln(v(0)) - \mu(0)) + \\ + \frac{1}{2}\sigma^2(1 - e^{-2\lambda \cdot s})). \quad (4)$$

We also give a formula describing the bounds $V_{L,R}(s)$ of the error corridor (e.g., within one standard deviation):

$$V_{L,R}(s) = \exp(\alpha \cdot s + \beta + \\ + e^{-\lambda \cdot s}\left(\ln(v(0)) - \mu(0)\right) \pm \sigma \cdot \sqrt{1 - e^{-2\lambda \cdot s}}), \quad (5)$$

where *R, L* are indices denoting, respectively, the right and left boundaries of the error corridor.

Formulas (2)–(5) are a direct consequence of the important property that the conditional distribution of the *OU* process is normal with parameters given by (1).

5. A practical example of the application of the method

The data on apartment sales in St. Petersburg published in issues 1483 through 1686 of the St. Petersburg Real Estate Bulletin covering the period from September 2011 through October 2015 are selected for the example. The St. Petersburg Real Estate Bulletin was published in print weekly through the end of 2019. The issues were selected on a one-issue-per-month basis, totaling 50 issues. Obviously, the full selection of a monthly issue contains mixed information about properties of different categories, and our example requires prices for properties of approximately the same categories. For this purpose, properties located in the Admiralteysky District of St. Petersburg were selected from citywide information. The Admiralteysky District is characterized by extensive "old stock" development, with relatively few premium properties in the central part of the city or old stock overlooking the great Neva

[4] Below we continue to use the already accepted designations, but in relation to the processed data.

River. Premium properties are also excluded from the sample. We are interested in properties that were in the journal for at least one period, so the stream consists of 48 samples (hereafter we will call them time slices), which are empirical samples of realizations of random variables $Y(j) = \ln(V(j)), j = \overline{0,47}$ and cover a period of 4 years. The standard deviations of all flow slices have a mean value of $\sigma = 0.22$ with insignificant changes over time (the standard deviation from its mean (0.22) is of order 0.01). Given the insignificant variations in the standard deviations, we will assume the standard deviation in the flow to be constant, equal to $\sigma = 0.22$ and make the prediction under this condition.

6. Spacing distribution

For each price value, the number of periods during which this price was maintained from slice to slice without a break was calculated. The resulting number is the length of the spacing τ corresponding to the given price, expressed in the total number of periods (one period – one month), taking integer values from 1 to 48. We then plot the empirical distribution of spacings by length (1, 2, 3, 4, ...), as well as their relative frequency to the total number of all spacings. The accumulated relative frequencies give the observed values of the empirical distribution function for spacing length. Our assumption is that the theoretical distribution of spacings obeys an exponential distribution law[5], that is, $F(t) = P(l \leqslant t) = 1 - e^{-\lambda \cdot t}$, where l is the random (theoretical) spacing length. We consider an additional probability distribution function $P(l > t) = 1 - F(t) = e^{-\lambda \cdot t}$. Obviously, $\ln(1 - F(t)) = -\lambda \cdot t$, i.e., the logarithm of the additional function depends linearly on t. Thus $\ln(1 - F(0)) = 0$. This property is used to fit the parameter λ of the exponential spacing distribution. *Figure 2* shows the observed values of the $1 - F(t)$ function at values $t = 0, 1, 2, 3, 4, 5, 6, 7, 8$ and their approximation by an exponent of the form $e^{-\lambda \cdot t}$.

The estimation of the parameter λ is obtained by the library function lm of the statistical package *R,* the value of $\lambda = 0.6510$.

7. Price distributions

Figure 3 shows the empirical distributions of prices of 1 sq. m. in the first six slices $V(j), j = \overline{0,5}$.

Figure 4 shows the *p*-value values of Kolmogorov–Smirnov tests for the correspondence of empirical distributions in 48 slices to the theoretical log-normal distribution, with the selected parameters.

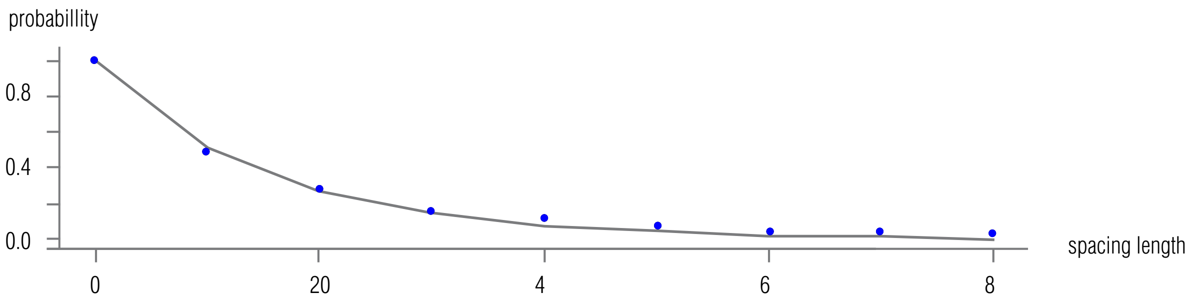


Fig. 2. Observed values of the additional function $1 - F(t)$ at values $t = 0, 1, 2, 3, 4, 5, 6, 7, 8$ (points) and their approximation by an exponent of the form $e^{-\lambda \cdot t}$.

[5] Since we consider slices at discrete time points, we observe geometrically distributed empirical spacings, which are the projection of continuous exponential spacings at discrete time points 0,1,2,3,... If the random variable is, $\xi \in \exp(\lambda)$, then $[\xi] \in Geom(p)$ (where square brackets denote the integer part of the number) and in our case, $p = 1 - e^{-\lambda}$.

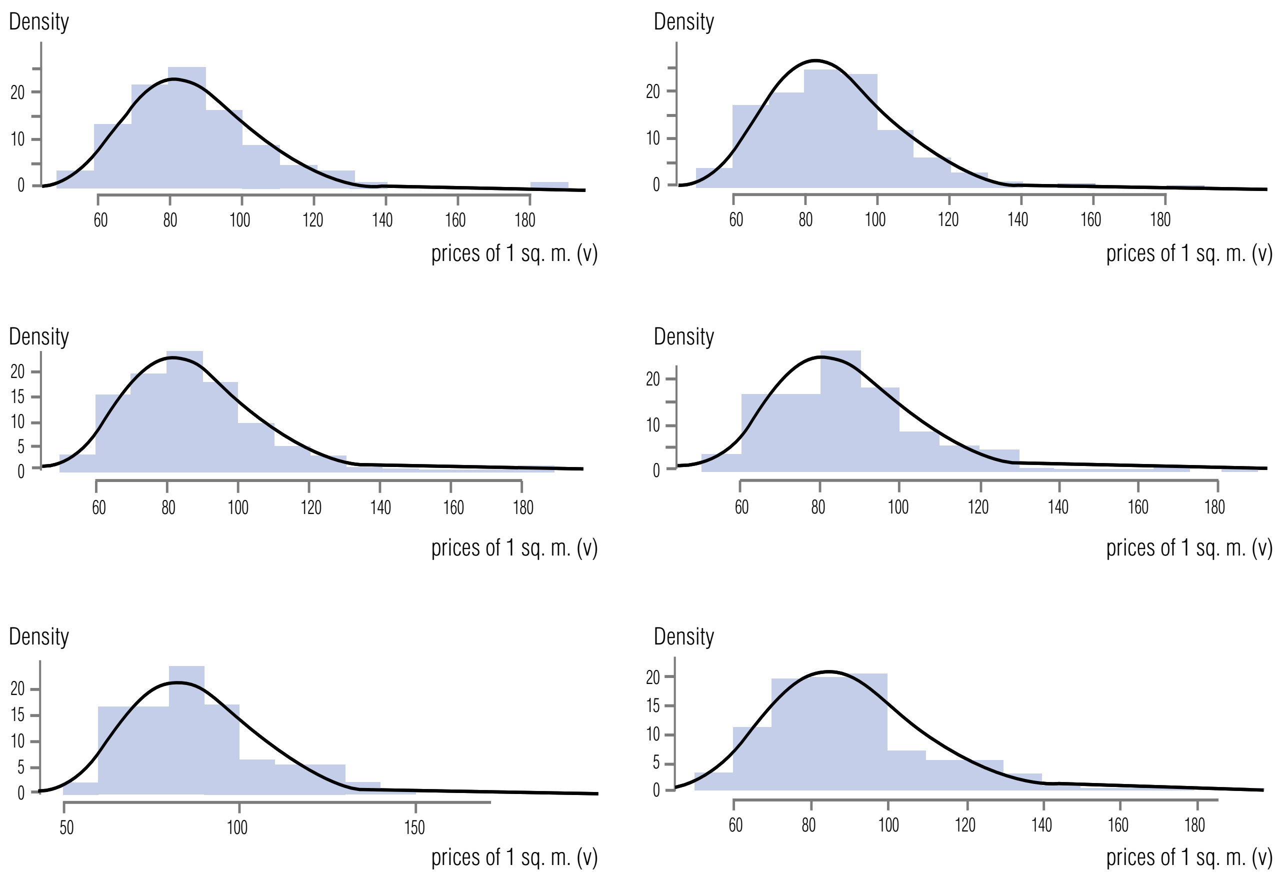
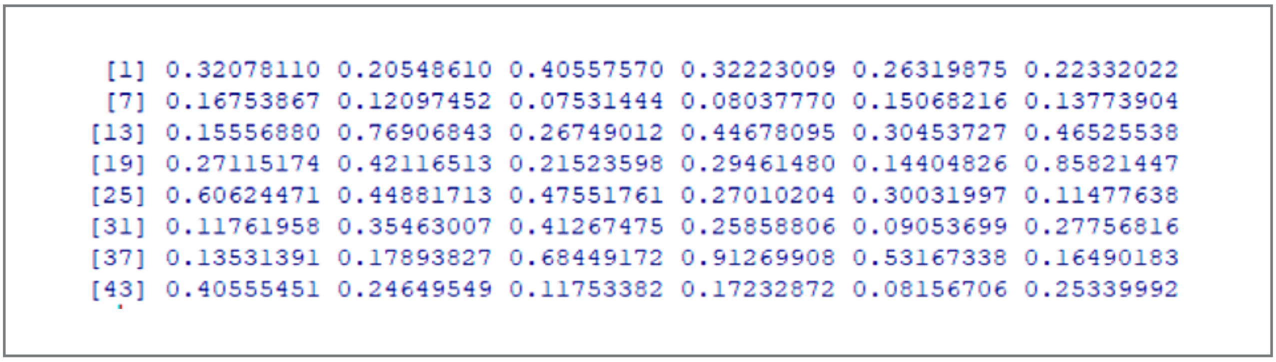


Fig. 3. Empirical distributions of prices of 1 sq. m. in the first six slices $V(j), j=\overline{1,6}$,
the line is the density of the model log-normal distribution.



```
 [1] 0.32078110 0.20548610 0.40557570 0.32223009 0.26319875 0.22332022
 [7] 0.16753867 0.12097452 0.07531444 0.08037770 0.15068216 0.13773904
[13] 0.15556880 0.76906843 0.26749012 0.44678095 0.30453727 0.46525538
[19] 0.27115174 0.42116513 0.21523598 0.29461480 0.14404826 0.85821447
[25] 0.60624471 0.44881713 0.47551761 0.27010204 0.30031997 0.11477638
[31] 0.11761958 0.35463007 0.41267475 0.25858806 0.09053699 0.27756816
[37] 0.13531391 0.17893827 0.68449172 0.91269908 0.53167338 0.16490183
[43] 0.40555451 0.24649549 0.11753382 0.17232872 0.08156706 0.25339992
```

*Fig. 4.*Screenshot of the ***R*** statistical package window with *p*-value values
of Kolmogorov–Smirnov tests for the correspondence of empirical distributions in 48 slices $V(j), j=\overline{0,47}$
of the theoretical log-normal distribution, with the selected parameters.

Thus, there is no reason to reject the hypothesis of log-normal distributions of prices in all 48 slices (or normal distributions of logarithms of prices).

8. Independence of copies of the random sequence

Let's consider the independence of copies of a random sequence $Y(j) = \ln(V(j)), j = \overline{0,47}$. We first note that the number of copies in each slice is different and varies from 254 to 729. It is known that the variance of the sum of two random variables is equal to the sum of the variance plus twice the covariance. We require evidence of uncorrelatedness of the records in the slices. To this end, we examine the behavior of the accumulated variance in each of the 48 slices $\ln(V(j))$, $j = \overline{0,47}$.

In the sample of each slice, the following steps are performed:

- a subsample is formed by randomly selecting a fixed number of elements from the sample (slice) (e.g. 20, the minimum sample size in the slice is 254, the maximum is 729), the variance is calculated;
- the next 20 elements are selected from the remaining elements of the sample, the variance is calculated, the result is added to the variance obtained at the previous step,
- the procedure continues until the sample is exhausted (the number of such steps is marked on the horizontal axis in *Fig. 5*).

Then the dependence of the variance accumulated in this way on the number of steps is considered. The linear character of the accumulated variance indicates the absence of a correlation term in the obtained sums. Since in our case the summarized quantities obey the normal distribution law, the non-correlation indicates independence.

Figure 5 shows the behavior of accumulated variance in the first slice, the behavior of accumulated variance in the first slice when the random selection procedure is repeated 1000 times, and the behavior of accumulated variance in all other slices.

9. Forecast construction

To construct the forecast, we should identify the linear trend in the data, estimate the PSI-process parameter λ, and check the joint normality of the mean values, which will provide confirmation of the Markov property of the observed process[6] . The standard deviation of the process is assumed above to be constant and equal to $\sigma = 0.22$.

Figure 6 shows a series of mean values of logarithms $Y(j) = \ln(V(j), j = \overline{0,47}$.

The linear trend is estimated using the library function lm of the statistical package *R*. We remove the linear trend, normalize the data by the standard deviation $\sigma = 0.22$, and average the data in each slice.

We obtain a series (*Fig. 7*) of average values of the process $\hat{Y}(j), j = \overline{0,47}$. The graph visually corresponds to the trajectory graph of the stationary process.

Figure 8 shows a plot of the autocorrelation function for a series of mean values of the process $\hat{Y}(j), j = \overline{0,47}$, with lags up to 7 and its approximation of the exponential of the form $e^{-\lambda \cdot s}$.

The estimation of the parameter $\lambda = 0.6502$, is obtained by applying the library function lm of the statistical package *R*.

Figure 9 shows the plot of the partial autocorrelation function for a series of mean values of the process $\hat{Y}(j), j = \overline{0,47}$, with lags up to 7.

---

[6] The Markov property is also indicated by the type of pacf in *Figure 9*, where there is exactly one significant peak per unit.

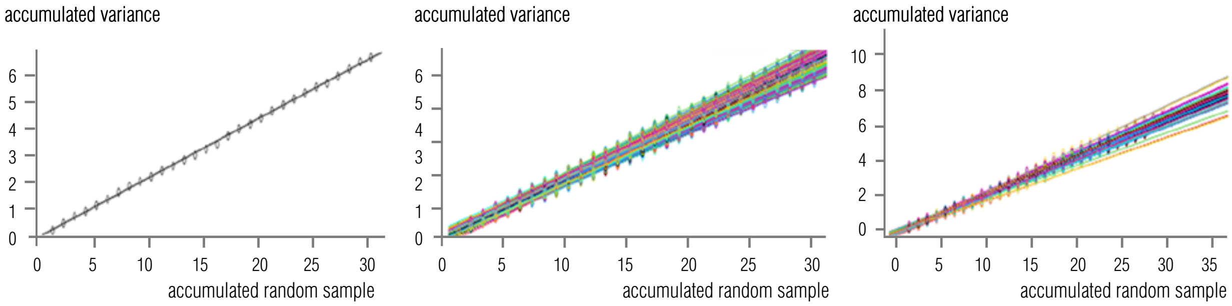


Fig 5. Left figure – accumulated variance of slice $\ln(V(1))$,
center figure – accumulated variance of slice $\ln(V(1))$
at 1000 repetitions of the random subsample selection procedure,
right figure – accumulated variance of all slice $\ln(V(j)), j = \overline{0, 47}$.

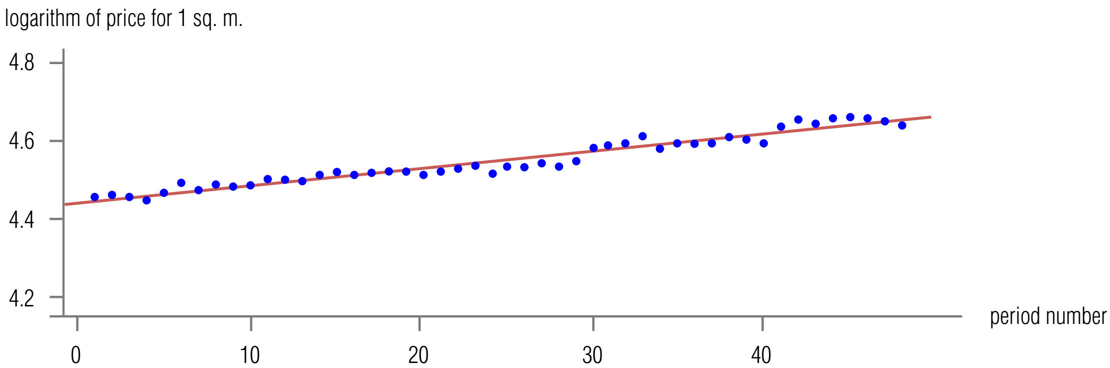


Fig. 6. Series of mean values for $Y(j) = \ln(V(j)), \ \ j = \overline{0, 47}$,
trend line equation $4.443 + 0.00432 \cdot t$, where t is the time in periods (1 period = 1 month)
from the beginning of observations (with j = 1, the value of t = 1/4).

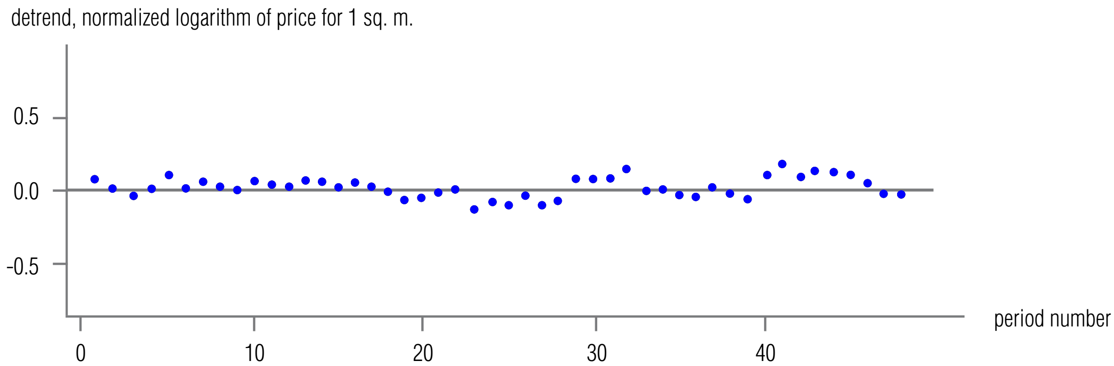


Fig. 7. A series of mean values of the process $\hat{Y}(j), j = \overline{0, 47}$.

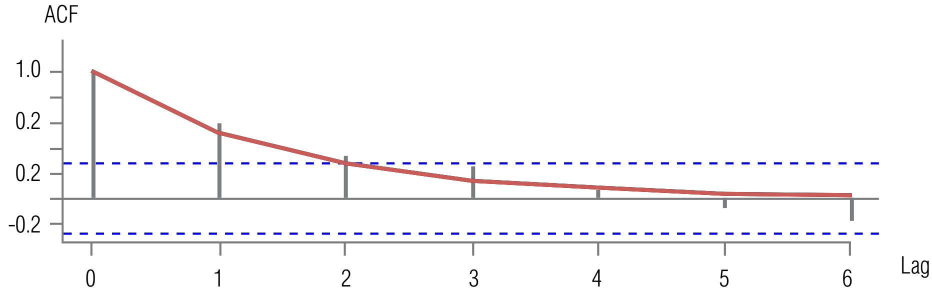


Fig. 8. Graph of the autocorrelation function for a series of mean values of the process $\hat{Y}(j), j=\overline{0, 47}$, with lags up to 7 and its approximation by an exponent of the form $e^{-0,6502 \cdot s}$.

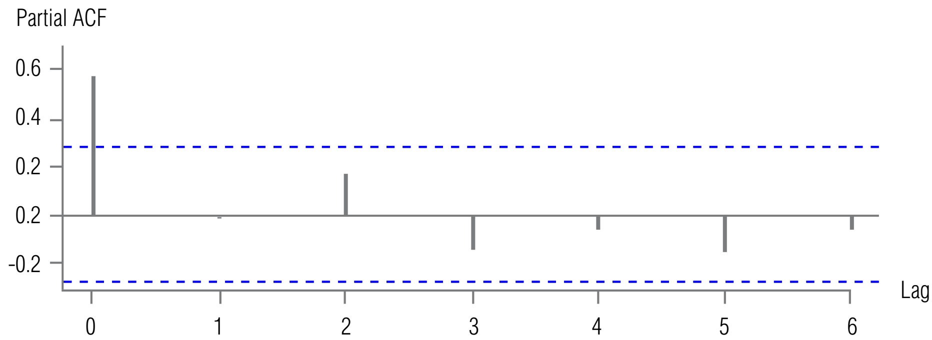


Fig. 9. Plot of the partial autocorrelation function for a series of process mean values $\hat{Y}(j), j=\overline{0, 47}$, with lags up to 7.

10. Checking the joint normality of the mean values of $\hat{Y}(j), j=\overline{0, 47}$

In the case of continuously substituted objects and prices, different sample sizes in each slice, checking joint normality by library tests (such as, for example, the MVN package tests [38, 39]) presents significant difficulties. To test for joint normality of the normalized sums of the mean of our PSI-process, we use the form of the covariance function and the fact that the normalized sums of the PSI-processes converge to the Ornstein–Uhlenbeck process. Here we rely on the representation of the Ornstein–Uhlenbeck process with discrete time as a first order autoregression with Gaussian white noise.

Let ε_0, ε_1, ε_2, ... be independent identically distributed standard normal random variables ($\varepsilon_j \in N(0, 1)$, $j=\overline{0, +\infty}$). Let us define the Ornstein–Uhlenbeck process with discrete time and speed $\lambda > 0$ in recurrent form:

$$\begin{cases} u_0 = \varepsilon_0 \\ u_{n+1} = e^{-\lambda} u_n + \sqrt{1-e^{-2\lambda}}\, \varepsilon_{n+1}. \end{cases}$$

Obviously $u_j \in N(0, 1), j=\overline{0, +\infty}$. You can write down

$$\begin{cases} \varepsilon_0 = u_0 \\ \varepsilon_{n+1} = \dfrac{u_{n+1} - e^{-\lambda} u_n}{\sqrt{1-e^{-2\lambda}}}. \end{cases} \quad (6)$$

We denote the number of elements in samples at each moment of discrete time as $N(j), j=\overline{0, +\infty}$ and set

$$u_j = \frac{1}{\sqrt{N(j)}} \sum_{i=1}^{N(j)} \hat{Y}_i(j), \quad j=\overline{0, +\infty}. \quad (7)$$

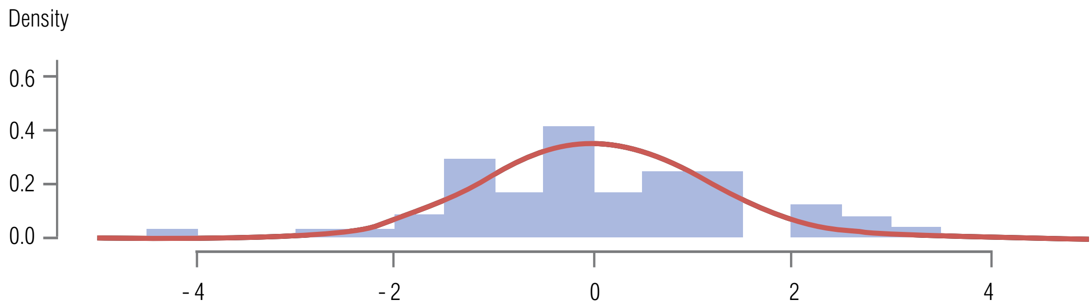


Fig. 10. Empirical distribution of values $\hat{\varepsilon}_j, j = \overline{0, 47}$ and its approximation by the normal distribution density.
Model distribution parameters $\mu = 0.0267$, $\sigma = 1.1$.

Applying formulas (6) to quantities (7), we obtain estimates $\hat{\varepsilon}_0, \hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots$ for $\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots$. From the independence and normality of the values $\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots$ the joint normality of values (7) follows.

The data consists of 48 samples (each of different lengths) corresponding to 48 discrete points in time. For each of them we set $\hat{Y}(j)$, $j = \overline{0, 47}$, normalized sums of averages (7), estimates $\hat{\varepsilon}_0, \hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots$ (6). To do this, we need to identify the trend (shown above), estimate the parameter λ (we obtained it twice, $\lambda \approx 0.65$), the standard deviation was previously assumed to be constant $\sigma = 0.22$. *Figure 10* shows the empirical distribution of the values $\hat{\varepsilon}_j, j = \overline{0, 47}$ and its approximation by the normal distribution density. *Figure 11* shows a screenshot of the results of statistical tests for the normality of values $\hat{\varepsilon}_j$, $j = \overline{0, 47}$ using the Kolmogorov–Smirnov, Shapiro, Andersen–Darling tests.

```
> ks.test(eps,"pnorm",mean=0.0267,sd=1.1)

        Exact one-sample Kolmogorov-Smirnov test

data:  eps
D = 0.096711, p-value = 0.7239
alternative hypothesis: two-sided

> shapiro.test(eps)

        Shapiro-Wilk normality test

data:  eps
W = 0.97975, p-value = 0.5684

> ad.test(eps)

        Anderson-Darling normality test

data:  eps
A = 0.25751, p-value = 0.7054
```

Fig. 11. Screenshot of the results of statistical tests of Kolmogorov–Smirnov, Shapiro, Andersen–Darling. The test results confirm the noise normality hypothesis (6).

Thus, the verifications carried out allow us to confirm the following hypotheses:

1) spacing (τ_i), $i = \overline{1, 48}$ obey an exponential distribution law with parameter $\lambda \approx 0.65$;
2) each component of the discrete process $\hat{Y}(j)$, $j = \overline{0, 47}$ is standard normal;
3) the observed copies of the random series $\ln(V(j))$ are independent (i is the copy number, j is the slice number);
4) the autocorrelation function can be approximated (*Fig. 8*) by an exponential function $e^{-\lambda \cdot s}$, with a parameter $\lambda \approx 0.65$, thus the correlation coefficient for a 6-step forecast (that is, six months ahead) can be set as $e^{-0.65 \cdot 6} = e^{-3.9}$;
5) the form of the partial autocorrelation function presented in *Fig. 9* indicates that $\hat{Y}(j)$ is conditionally uncorrelated with $\hat{Y}(k)$, $|k - j| > 1$, $j = \overline{0, 47}$, provided that the values of $\hat{Y}$ at moments of time strictly between k and i;
6) the Markov property of the discrete process $\hat{Y}(j)$, $j = \overline{0, 47}$ follows from the joint normality and the

form of the covariance function. The Markov property is also confirmed by the form of the partial autocovariance function.

11. Forecast

The above checks allow us to establish the correspondence of the observed data to the PSI-process model. The Markov property of the observed discrete process $\hat{Y}(j), j = \overline{0,47}$ allows us to make a forecast only based on the latest distribution of the random variable V (the cost of 1 sq.m. of secondary residential real estate in the Admiralteysky district, in the 48th period corresponding to September 2015). The empirical distribution of the random variable $V(48)$ is satisfactorily approximated logarithmically normal distribution with parameters $\mu(0) = 4.64$, $\sigma = 0.22$ (p-value of the Kolmogorov–Smirnov test is 0.2534, see *Fig. 4*).

The mathematical expectation of the model distribution in the 48th period is equal to $e^{4.64+0.5\cdot 0.22^2} = 106.081$ thousand rubles for 1 sq. m. The median of the model distribution in the 48th period is equal to $e^{4.64} = 103.544$ thousand rubles for 1 sq. m. The mode of the model distribution in the 48th period is equal to $e^{4.64-0.22^2} = 98.652$ thousand rubles for 1 sq. m.

Note that the average price exceeds the most probable value by 7.5%, the minimum price is 1 sq. m. in the empirical sample in the 48th period 63.265 thousand rubles; maximum price 1 sq. m. in the empirical sample in the 48th period 172.556 thousand rubles.

The correlation function of the normalized sums of the discrete process $\hat{Y}(j), j = \overline{0,47}$ has an exponential form with the parameter $\lambda = 0.65$. Moreover, we obtained this value twice: as a parameter of the exponential approximating the correlation function, and as a parameter of the exponential distribution of spacing).

The correlation coefficient for the logarithm of a pair of arbitrary sections of the process $V(t)$, forming a two-dimensional random vector $(V(0), V(s))$, is equal to $\rho(\ln(V(0)), \ln(V(s))) = e^{-0.65\cdot s}$. Formulas (2)–(4) have the form:

$$Mode(V(s)|V(0) = v(0)) =$$
$$= \exp(0.00432 \cdot s + 4.443 + e^{-0.65\cdot s}(\ln(v(0)) - \mu(0) -$$
$$- 0.22^2(1 - e^{-2\cdot 0.65\cdot s})),$$

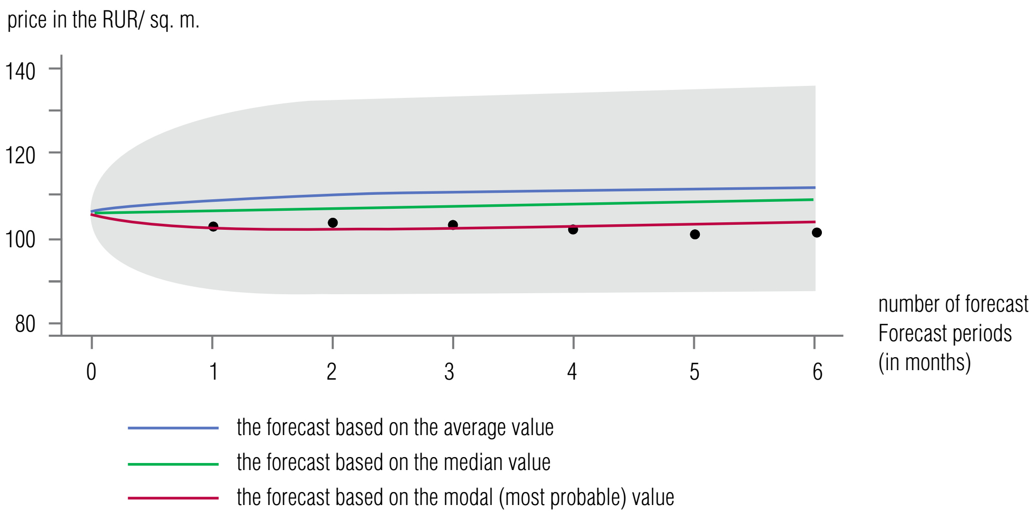


Fig. 12. Forecast for the price value $v(0) = 106.081$ thousand rubles per 1 sq. m. – the average price in the last observation period. Gray background – corridor within one standard deviation.

$$Median(V(s)|V(0) = v(0)) =$$
$$= \exp(0.00432 \cdot s + 4.443 + e^{-0.65 \cdot s}(\ln(v(0)) - \mu(0))),$$

$$E(V(s)|V(0) = v(0)) =$$
$$= \exp(0.00432 \cdot s + 4.443 + e^{-0,65 \cdot s}(\ln(v(0)) - \mu(0)) +$$
$$+ \frac{1}{2} 0.22^2 (1 - e^{-2 \cdot 0.65 \cdot s})).$$

Formula (5), which describes the boundaries of the $V_{L,R}(s)$ error corridor (for example, within one standard deviation), takes the form:

$$V_{L,R}(s) = \exp\Big(0.00432 \cdot t + 4.443 +$$
$$+ e^{-0\ 65 \cdot s}\big(\ln(v(0)) - \mu(0)\big) \pm 0.22 \cdot \sqrt{1 - e^{-2 \cdot 0.65 \cdot s}}\Big).$$

For any property, price is 1 sq. m. which $V(0) = v(0)$ was set in the last period (as stated above, corresponds to September 2015), a forecast can be made for the next 6 periods.

Let us build a forecast for 6 periods (six months in advance) for a property that in the last period had a price of 106.081 thousand rubles for 1 sq. m. (average price for slice number 47).

Figure 12 shows the forecast for the price value $v(0) = 106.081$ thousand rubles per 1 sq. m.

In *Fig. 12*, dots mark the actual average prices for 1 sq. m. in the same set of real estate properties (mass market, Admiralteysky district), which are:

in October 2015 103.100 thousand rubles for 1 sq. m.,

in November 2015 103.492 thousand rubles for 1 sq. m.,

in December 2015 103.247 thousand rubles for 1 sq. m.,

in January 2016 102.027 thousand rubles for 1 sq. m.,

in February 2016 101.044 thousand rubles for 1 sq. m.,

in March 2016 102.046 thousand rubles for 1 sq. m.

Conclusion

1. The complex of checks carried out made it possible to establish that the studied data follows the PSI-process model.
2. *Figure 11* shows the forecast for a price equal to the average value in the last observation period. The black dots show the actually observed average price values of 1 sq. m. in the next 6 months, they are located near the modal forecast line, which serves as a good verification of the forecast based on the most probable value. The line of modal values indicates the most probable value of the future price, provided that the price in the last slice was equal to the average value. This is a consequence of the characteristic of the lognormal distribution of prices and the fact that the average values of the logarithms of prices follow the Ornstein–Uhlenbeck process.
3. The parameter $\lambda = 0.65$ indicates, among other things, the average time during which the price of a property in the observed market sector jumps; it is $1/\lambda = 1.54$ periods. In the example presented, the period is 1 month.
4. The main advantages of forecasting based on the characteristics of PSI-processes:

- the standard deviation of the forecast stabilizes over time at the level of constant variance of the random process, in contrast to moving average models that accumulate forecast error at each step;
- the ability to build a forecast not only for average values, but also for any object put up for sale in the last observation period at a certain price. ■

References

1. Friedman J., Ordway N. (1997) *Analysis and assessment of income-generating real estate*. M.: Delo (in Russian).
2. Sternik G.M., Sternik S.G. (2018) Methodology for forecasting the Russian real estate market, *Mechanization of construction*, no. 8(830), p. 57 (in Russian).

3. Sternik G.M., Sternik S.G. (2015) Residential real estate market in Moscow and the Moscow region: current state and price forecast. *Territory Development Management*, no. 4, pp. 31–36 (in Russian).

4. Sternik G.M., Pechenkina A.V. (2007) Forecast of supply prices for apartments on the Russian housing market (macroeconomic approach). *Property Relations in the Russian Federation*, no. 10(73), pp. 11–18 (in Russian).

5. Sternik G.M., Sternik S.G., Sviridov A.V. (2014) Development and improvement of forecasting methods in the residential real estate market. *Urban Studies and Real Estate Market*, no. 1, pp. 53–93 (in Russian).

6. Alekseeva M.O., Paidiganova M.Yu. (2019) Analysis and forecasting of prices for apartments in the Republic of Maria El. *Alleya Nauki*, vol. 1, no. 9(36), pp. 65–70 (in Russian).

7. Drobyshevsky S.M. (2009) Analysis of the possibility of a "bubble" in the Russian real estate market. *IEP Scientific Works*, no. 128 (in Russian).

8. Evstafiev A.I., Gordienko V.A. (2009) Forecasting indicators of the real estate market using neural networks. *University News. North-Caucasian Region. Social Sciences Series*, no. 5, pp.83–89 (in Russian).

9. Zenchik A.S., Morozova N.N. (2016) Forecasting real estate prices taking into account seasonality. *Youth Scientific Forum: Social and Economic Sciences*, no. 11(40), pp. 397–402 (in Russian).

10. Koshkin V.S., Boronina N.Yu. (2022) Forecasting the market value of commercial real estate based on indicators of economic development of the territory of Barnaul. *Implementation of priority programs for the development of the agro-industrial complex. Collection of scientific papers based on the results of the X International Scientific and Practical Conference dedicated to the memory of the Honored Scientist of the Russian Federation and Kabardino-Balkaria, Professor B.Kh. Zherukova, Nalchik*, pp. 312–315 (in Russian).

11. Molchanova M.Yu., Pechenkina A.V. (2011) Features of using methods of fundamental and technical analysis when forecasting prices on the regional real estate market. *Bulletin of Perm University. Series "Economics"*, no. 3(10), pp. 53–54 (in Russian).

12. Nikitina N.S. (2022) Forecasting the real estate price index taking into account seasonality, *Economic Development of Russia*, vol. 29, no. 6, pp. 23–28 (in Russian).

13. Nurmukhametov I.M. (2014) Analysis and forecasting of real estate prices in the Mari-El Republic. *Industrial development of Russia: problems, prospects. Proceedings of the XII International Scientific and Practical Conference of teachers, scientists, specialists, graduate students, students*, pp. 97–104 (in Russian).

14. Pechenkina A.V. (2010) Using a multi-level factor model when predicting the situation in the regional real estate market (using the example of the Perm Territory). *Property Relations in the Russian Federation*, no. 11(110), pp. 57–72 (in Russian).

15. Pitulin S.S. (2019) Construction of ARIMA models for analyzing the dynamics of real estate prices in the Smolensk region. *Internauka*, no. 20-1(102), pp. 63–67 (in Russian).

16. Rubinshtein E.D., Osipenko N.S. (2015) Analysis of the real estate market and its forecasting. *Theory and Practice of Social Development*, no. 12, pp. 140–143 (in Russian).

17. Clapp J.M., Giaccotto C. (2002) Evaluating house price forecasts. *Journal of Real Estate Research*, vol. 24, no. 1, pp. 1–25.

18. Crawford G.W., Fratantoni M.C. (2003) Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house price. *Real Estate Economics*, vol. 31, no. 2, pp. 223–243.

19. Geltner D.M., Miller N.G., Clayton J., Eichholtz P. (2006) *Commercial real estate analysis and investments*. South-Western Educational Pub.

20. Gotham K.F. (2006) The Secondary Circuit of Capital Reconsidered: Globalization and the U.S. Real Estate Sector. *American Journal of Sociology*, vol. 112, no. 1, p. 231–275. https://doi.org/10.1086/502695

21. Green R.K. (2008) Imperfect information and the housing finance crisis: A descriptive overview. *Journal of Housing Economics*, vol. 17, no. 4, p. 262–271. https://doi.org/10.1016/j.jhe.2008.09.003

22. Green R.K. (1997) Follow the leader: How changes in residential and nonresidential investment predict changes in GDP. *Real Estate Economics*, vol. 25, no. 2, p. 253–270.

23. Laskin M.B., Cherkesova P.A. (2020) Comparison of market and cadastral data for predicting the market value of real estate. *Statistics and Economics*, vol. 17, no. 4, pp. 44–54 (in Russian).

24. Fedorov N.I. (2020) Forecasting prices for residential real estate on the Chelyabinsk market using machine-learning methods. *Student and scientific and technological progress. Materials of the XLIV scientific conference of young scientists, Chelyabinsk*, pp. 223–226 (in Russian).

25. Li Y., Xiang Z., Xiong T. (2020) The behavioral mechanism and forecasting of Beijing housing prices from multiscale perspective. *Discrete Dynamics in Nature and Society*, vol. 2020, article ID 5375206. https://doi.org/10.1155/2020/5375206

26. Ohnishi T., Mizuno T., Shimizu C., Watanabe T. (2011) On the evolution of the house price distribution. *Columbia Business School. Center of Japanese Economy and Business. Working Paper Series*, no. 296, pp. 1–20.

27. Raymond Y.C.T. (1997) An application of the ARIMA model to real estate prices in Hong-Kong. *Journal of Property Finance*, vol. 8, no. 2, p. 152–163. https://doi.org/10.1108/09588689710167843

28. Singh Y. (2005) *Model for forecasting price of houses in city of Stillwater*, M.S. dissertation, Oklahoma State University.

29. Sa'at N.F., Maimun N.H.A., Idris N.H. (2021) Enhancing the accuracy of Malaysian house price forecasting: a comparative analyses on the forecasting performance between the hedonic price model and artificial neural network model. *Planning Malaysia*, vol. 19, no. 3, p. 249–259.

30. Rusakov O., Laskin M., Jaksumbaeva O. (2016) Pricing in the real estate market as a stochastic limit. Log Normal approximation. *International Journal of the Mathematical Models and Methods in Applied Sciences*, vol. 10, pp. 229–236.

31. Rusakov O., Laskin M., Jaksumbaeva O., Ivakina A. (2015) Pricing in real estate market as a stochastic limit. Lognormal approximation, *2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*. Institute of Electrical and Electronics Engineers Inc., pp. 235–239. https://doi.org/10.1109/MCSI.2015.48

32. Aitchinson J., Brown J.A.C. (1963) *The Lognormal distribution with special references to its uses in economics*. Cambridge: University Press.

33. Rusakov O.V. (2017) Pseudo-Poisson processes with stochastic intensity and the class of processes generalizing the Ornstein-Uhlenbeck process. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, vol. 4, no. 2, pp. 247–257 (in Russian).

34. Yakubovich Y., Rusakov O., Gushchin A. (2022) Functional limit theorem for the sums of PSI-processes with random intensities. *Mathematics*, vol. 10, no. 21, 3955. https://doi.org/10.3390/math10213955

35. Feller W. (1971) *An introduction to probability theory and its applications*. Wiley.

36. Laskin M.B. (2017) Adjustment of market value according to the pricing factor "object area". *Property Relations in the Russian Federation*, no. 8(191), pp. 86–99 (in Russian).

37. Laskin M.B. (2018) Determination of the trading discount based on market data and cadastral value. *Business Informatics*, no. 3(45), pp. 53–61. https://doi.org/10.17323/1998-0663.2018.3.53.61

38. Korkmaz S., Goksuluk D., Zararsiz G. (2014) MVN: An R package for assessing multivariate normality. *The R Journal*, vol. 6/2, pp. 151–162.

39. Korkmaz S., Goksuluk D., Zararsiz G. (2022) *Package 'MVN'*. Available at: https://cran.r-project.org/web/packages/MVN/MVN.pdf (accessed 30 November 2023).

About the authors

Mikhail B. Laskin

Cand. Sci. (Phys.-Math.), Associate Professor;

Senior Researcher, Laboratory of Intelligent Systems, Saint Petersburg Federal Research Center, Russian Academy of Sciences, 39, 14th Line, Vasilyevsky Island, St. Petersburg 199178, Russia;

E-mail: laskinmb@yahoo.com

ORCID: 0000-0002-0143-4164

Oleg V. Rusakov

Cand. Sci. (Phys.-Math.);

Associate Professor, Department of Probability Theory and Mathematical Statistics, Saint Petersburg State University, 7/9, Universitetskaya Nab., St. Petersburg 199034, Russia;

E-mail: ovirusakov@yahoo.co.uk

ORCID: 0000-0001-7444-9896

DOI: 10.17323/2587-814X.2023.4.25.40

The impact of economic complexity and industry specialization on the gross regional product of Russian regions

Alexander V. Kudrov
E-mail: kovlal@inbox.ru

Central Economic and Mathematical Institute, Russian Academy of Sciences
Address: 47, Nakhimovsky Prospect, Moscow 117418, Russia

Abstract

The economic complexity index defines the basis of the modern theory of economic complexity and reflects the level of knowledge embedded in the production structure of the economy. This study examines the direct relationship between the economic complexity index and gross regional product (GRP) while taking into account other factors of the GRP production function in its generalized representation. As a result, we can isolate the impact of the economic complexity index from other phenomena. The non-linear nature of the relationship between economic complexity and GRP is revealed, and the direct relationship is manifested only at sufficiently high values of economic complexity, exceeding a certain threshold, which is found endogenously using econometric methods. In addition, the paper studies the relationship between economic complexity and indices of sectoral specialization. We found that there is a direct relationship between economic complexity and the extractive industry index and no relationship with the level of development of manufacturing industry. We obtained a clarification of the generalized production function of GRP, in which the threshold effect of the influence of economic complexity manifested itself as a factor of nonlinear dependence describing the elasticity of labor: a high level of economic complexity provides greater labor productivity. Overall, the results of the study of the dependence of GRP on economic complexity lead to the conclusion that increasing economic complexity can be an effective way to stimulate economic growth and development, but only starting from a certain threshold level. This suggests that an economy must reach a minimum level of diversity and complexity in its industrial activities before it can experience the productivity gains necessary for substantial GRP growth.

Keywords: economic complexity index, sectoral specialization, generalized production function, direct relationships, nonparametric regression, nonlinear regression, returns to scale

Citation: Kudrov A.V. (2023) The impact of economic complexity and industry specialization on the gross regional product of Russian regions. *Business Informatics*, vol. 17, no. 4, pp. 25–40.
DOI: 10.17323/2587-814X.2023.4.25.40

Introduction

For a relatively long time, economists have agreed that a country's ability to create and distribute income depends on its productive structure, as discussed in papers such as [1–3]. Paper [4] presents a study of the convergence of productivity levels across US states and finds that higher productivity levels are associated with a more diverse and complex production structure.

However, quantifying the productive structure of a country is difficult. Various approaches have been attempted, such as the concentration index, which measures the share of agriculture, manufacturing or services in the economy, as well as aggregate measures of diversity and concentration [5]. Other approaches measure diversification by dividing sectors into related and unrelated sectors [6–8]. However, these methods have their limitations, including the possibility of some bias, since large countries or regions tend to be more diversified. In addition, they do not take into account the interconnections between different economic activities, complexity and sophistication of production activities.

These shortcomings are resolved by considering the identified comparative advantages and constructing an economic complexity index [9–11]. The economic complexity index allows us to obtain estimates of the complexity of economic structures, taking into account both the diversity and uniqueness of sectors. This allows us to reflect both the breadth and depth of the economic structure.

One of the most important aspects of economic complexity is sectoral network structure, which measures the degree to which different sectors of the economy are interconnected through production processes. This interconnectedness is believed to facilitate the transfer of knowledge, technology and other resources between sectors and support economic growth. Large values of the economic complexity index indicate that the structure of the economy is dominated by interconnected sectors. For example, sectors with long production cycles, such as electronics and engineering, require higher levels of coordination and knowledge and therefore have high levels of economic complexity. In contrast, economic structures dominated by primary and agricultural sectors have low values of economic complexity.

The paper [12] presents calculations of the economic complexity index for countries and shows how it can be used to forecast economic growth and identify potential areas for diversification and development of the economies of countries.

The relationship between economic complexity and gross domestic product (GDP) is of great interest to economists, since GDP is a widely accepted indicator of regional production and economic development.

The authors of [11] have shown that countries with a more complex production structure tend to have higher levels of economic growth and higher GDP per capita, which in turn are associated with lower poverty rates and better social welfare [13]. Hence, one may conclude that development policies should aim to create conditions that will stimulate growth in economic sophistication (for a more detailed discussion see [10]).

In recent years, statistical studies have used the economic complexity index as an explanatory factor for economic growth, knowledge level, human capital, inequality and other socio-economic indicators [12, 14, 15]. However, the relationship between economic complexity and socioeconomic indicators is not always unambiguous, and there are other factors that may influence this relationship. As will be shown in this paper, often the assumption or conclusions that there is a direct relationship of GRP with economic complexity is erroneous, because, as a rule, other basic indicators of the economy and science are not considered.

In [16], a generalized GRP production function was obtained, in which regional output depends on the number of employed persons (L), the value of fixed assets (K) with their elasticity coefficients, which are given by the sectoral structure of GRP, and the number of researchers (P) (distinguished as an additional production factor with a constant elasticity coefficient). These factors of the production function will be considered as the main characteristics of the economy.

The purpose of this paper is to test two hypotheses using data on the regions of the Russian Federation. First, we will examine whether there is a direct relationship between the index of economic complexity and GRP. Second, we will examine whether there is a direct relationship between economic complexity and the factors of the generalized production function. For this purpose, we will use the methodology of finding direct relationships and hypothesis family testing [16, 17].

1. Data

Let us consider GRP for the year 2019 and the main factors of the generalized GRP production function from [16]:

- gross regional product for the year 2019 [18];
- fixed assets at the end of 2019 [18];
- average annual number of employed persons for 2019 [18];
- indices of extractive (S_1) and manufacturing (S_2) industries for 2019 (see below);
- number of researchers in 2019 [18].

Let us explain in more detail about the indices of extractive (S_1) and manufacturing (S_2) industries, which characterize the sectoral features of the region. These indices were constructed by the author based on the data of GRP sectoral structure using component analysis with rotation and reflect the sectoral specialization of the regions under consideration (*Fig. 1*). The data of the GRP sectoral structure included the main six industries that determine the nature of the economy of the Russian regions [18]:

- agriculture, forestry, hunting, fishing and fish farming;
- mining, oil and gas;
- manufacturing;
- wholesale and retail;
- real estate operations;
- public administration and military security; social security.

Further, these indicators were expressed through two factors by the method of principal components with rotation (*Fig. 1*), which account for more than 80% of the explained variance in the data for the year 2019. For other close years, very similar results are observed, indicating a very slow change in the structure of regional GRP.

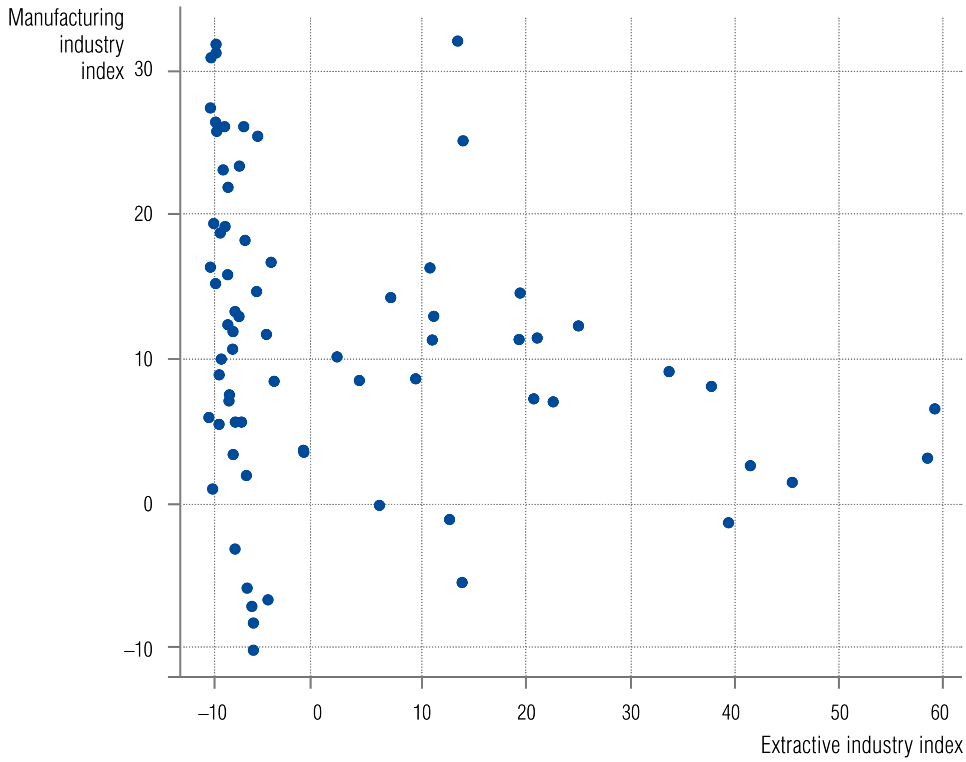


Fig. 1. Distribution of Russian regions in the space of sectoral orientation indices (listed in ascending order; based on data for 2019).

At the same time, we considered the data on tax revenues for 82 sectors of the Russian Federation regions [19], which reflect the production volumes of each sector of the economy for export and for domestic consumption. Based on these data, an index of economic complexity was constructed for 2019–2020 [20]. *Figure 2* shows the estimates we obtained of the economic complexity index. Note that the index of economic complexity, in fact, is equivalent to the generalized eigenvector of the matrix "region–region," the elements of which characterize the nested structures of economies.

The region of economic complexity values can be roughly divided into ranges, within which locally the points are well approximated by linear rank dependencies:

- **Range-1:** regions with a predominance of unique sectors in the economic structure. As a rule, these regions are characterized by specialization in the extractive industry. For them, the average value of the extractive industry index (+13.64). A sufficiently high average value of the manufacturing industry index (+11.05) indicates the presence of regions with a mixed-type structure, where manufacturing sectors are also sufficiently represented.
- **Range-2:** regions with a weakly diversified mix of strong sectors and non-unique sectors. These regions include regions with emerging economies. Average value of the extractive industry index (+7.16), average value of the manufacturing industry index (+7.45).
- **Range-3:** regions with highly diversified structures of strong sectors and long value chains. This includes regions characterized by the presence of long value chains and specialization in manufacturing. The average value of the extractive industry index (–3.87) indicates the absence of extractive industry

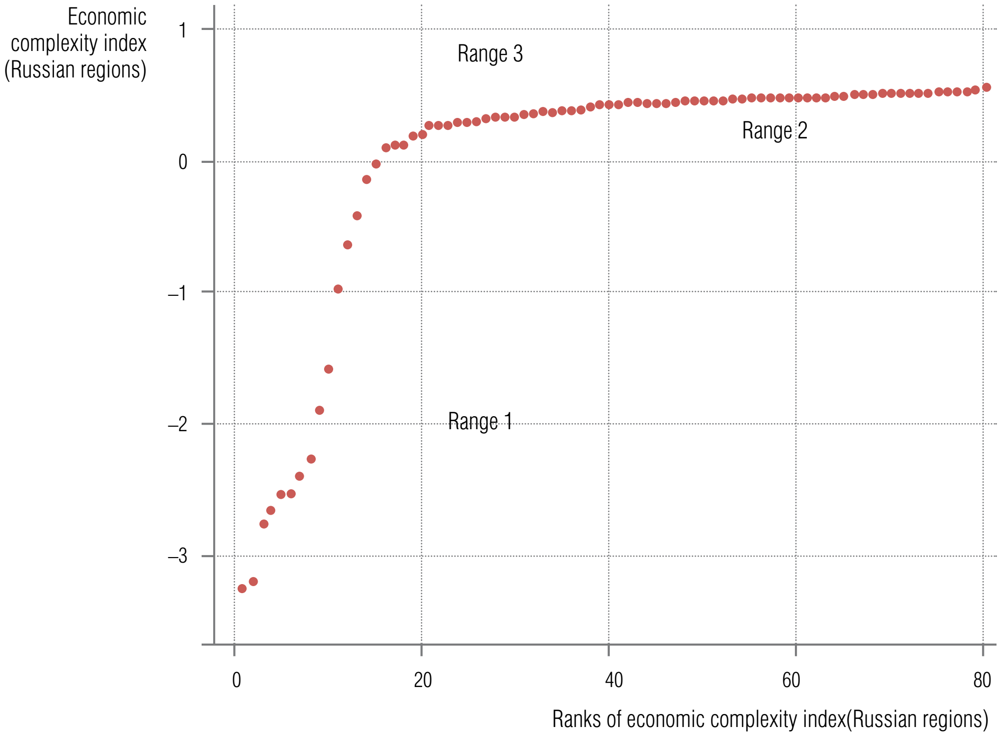


Fig. 2. Estimates of economic complexity by regions of the Russian Federation (listed by ascending order; based on data for 2019).

in most of these regions. The average value of the manufacturing industry index is (+17.68).

It should be noted that the smallest dispersion of economic complexity values is observed for points from Range 1, and the dispersion of points from Range 3 is the largest. The minimum average value of the extractive industry index is for regions with economic complexity in Range 1, and the maximum for Range 3.

2. Research methodology

The proposed methodology aims to obtain an analytical expression describing the impact of economic complexity on gross regional product (GRP). The methodology consists of several steps:

1. Identification of explanatory variables directly related to GRP. The first step is to identify the variables that are directly related to GRP (hereinafter referred to as Y_i – GRP of the *i*-th region). This is done using the technique of the so-called "causal analysis" or analysis of the direct relationships structure [16]. Let us explain this concept. If in a set of random variables (including both resultant and explanatory variables) $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_n)$, the conditional distribution of value Z_i from all others is determined *only by their part* $Z_j, Z_k, ..., Z_l$ (those not included in the condition can take any values). Let us denote by lowercase letters $z_1, ..., z_n$ the realized values of the corresponding random variables $Z_1, Z_2, ..., Z_n$. Then the definition of direct relationships can be written as:

$$P(z_i \mid z_1, ..., z_n) = P(z_i \mid z_j, z_k, ..., z_l) \;\; z_1, ..., z_n), \quad (1)$$

for all $z_1, ..., z_n$,

and the variables $Z_j, Z_k, ..., Z_l$ are called *directly related* to the variable Z_i. In the continuous case, the partial correlations of Z_i with the directly related (and only with them!) are not zero. Namely, if Z_i and Z_j are directly correlated, then different from zero will be:

$$corr\left(Z_i, Z_j | \boldsymbol{Z}_{-(i,j)}\right) = \\ = corr\left(Z_i - Pr_{Z_i}\left(\boldsymbol{Z}_{-(i,j)}\right), Z_j - Pr_{Zj}\left(\boldsymbol{Z}_{-(i,j)}\right)\right), \qquad (2)$$

where $\boldsymbol{Z}_{-(i,j)}$ – the set of variables excluding Z_i and Z_j;

$Pr_{Z_i}\left(\boldsymbol{Z}_{-(i,j)}\right)$ and $Pr_{Zj}\left(\boldsymbol{Z}_{-(i,j)}\right)$ – projection of Z_i and Z_j onto the linear subspace $sp\left(\boldsymbol{Z}_{-(i,j)}\right)$

2. Identification of the form of the relationship between GRP and economic complexity: monotonic or non-monotonic? To identify the non-monotonic relationship between GRP and economic complexity, the estimation of the non-parametric Nadaraya–Watson kernel regression $g_\tau(x)$ is used [21]:

$$g_\tau(x) = \frac{\sum_{i=1}^{N} w_i(x) \log(Y_i)}{\sum_{i=1}^{N} w_i(x)}, \qquad (3)$$

where $w_i(x) = k\left(\frac{x - X_i}{\pi}\right)$, $k(y) = \frac{\exp\left(-\frac{y^2}{2}\right)}{\sqrt{2\pi}}$;

$\log(Y_i)$ – logarithm of GRP for i-th region;

X_i – ranks of economic complexity for i-th region;

$k(y)$ – kernel of nonparametric regression (3) with parameter τ, τ– window width in nonparametric kernel regression (3).

Note that the window width τ was estimated using the so-called leave-one-out estimate cross validation, see [22] for details:

$$\tau_{opt} = \underset{\tau}{\operatorname{argmin}} \sum_{i=1}^{N} \left(\log\left(Y_i\right) - g_{\tau,(i)}\left(X_i\right)\right)^2, \qquad (4)$$

where (i) means that point i is not considered when computing the nonparametric estimate at point X_i. The use of cross validation with one point left out is particularly useful when the data size is small, as it allows the model to be trained on almost the entire data set. However, for large data, this cross-validation approach can be computationally expensive because the model has to be retrained for each individual data point.

3. Construction of nonlinear regression dependence of GRP on directly related explanatory variables. After the variables that are directly related to GRP (denoted by Y_i – GRP of the i-th region, $i = 1, ..., N$) are identified, the form of non-monotonic dependence of GRP and economic complexity is determined, a nonlinear regression dependence on these variables is constructed:

$$Y_i = f(x_i, \theta^*) + \varepsilon_i, \; i = 1, ..., N, \qquad (5)$$

where $f: \mathbb{R}^k \to \mathbb{R}$ a nonlinear function of the explanatory and directly related to Y_i variables, $x_i \in \mathbb{R}^k$;

$\boldsymbol{\theta}^* \in \mathbb{R}^p$ – vector of true values of the unknown parameters;

(ε_i) are assumed to be independent identically distributed random variables (not necessarily normally distributed) with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$.

Under the assumption that the function $f(\cdot)$ is known, the parameter vector $\boldsymbol{\theta}$ of model (5) is estimated as the solution to the following problem:

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \left(Y_i - f\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right)\right)^2. \qquad (6)$$

Finding a solution to this problem is done by numerical methods using the Levenberg–Marquardt algorithm [23, 24].

According to the results presented in [23, 24], under sufficiently large n and appropriate regularity assumptions (such as twice continuously differentiable $f(\boldsymbol{x}_i, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$), the LSE-estimator $\hat{\boldsymbol{\theta}}$ has an asymptotically normal distribution:

$$\hat{\boldsymbol{\theta}} \sim N_p\left(\boldsymbol{\theta}^*, \sigma^2\left[\left(F\left(\boldsymbol{\theta}^*\right)\right)^T F\left(\boldsymbol{\theta}^*\right)\right]^{-1}\right), \qquad (7)$$

where $F\left(\boldsymbol{\theta}^*\right) = \left[\left.\frac{\partial f\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]_{i,j} \in \mathbb{R}^{N \times p}$.

Thus, this methodology combines several statistical techniques including causal analysis, non-parametric estimation and non-linear regression to establish the relationship between economic complexity and GRP.

3. Research results

Let us analyze the mutual relations between economic complexity and the above-mentioned characteristics of science, the economy in the regions of the Russian Federation. For this purpose, the matrix of partial correlations was estimated using the data for the year 2019 and the hypotheses about the absence of direct relationship between each variable and economic complexity were consistently tested (*Table 1*).

Table 1 (right part) presents the results of testing the family of hypotheses about equality of partial correlations to zero; for a more detailed description of the testing procedure of the hypotheses considered [16]. Units indicate cases when there are no direct relationships between economic complexity and the corresponding variable.

As can be seen from *Table 1*, economic complexity is not related to the index of manufacturing industry, but it is related to the index of extractive industry. *Figure 3* provides visual confirmation for that.

Among all the variables considered, economic complexity has a statistically significant direct relationship with the extractive industry index. The existence of this relationship indicates that the scenario of a transition from an extractive-based economy (e.g., mining or oil, gas) to a more diversified one (in particular, oriented towards long value chains) is associated with an increase in the level of economic complexity.

The lack of a relationship between the economic complexity index and the manufacturing index may imply that the mere presence of manufacturing in the economy is not sufficient to increase its complexity. This may be the case if manufacturing is concentrated in a few low-complexity industries or if other sectors of the economy remain underdeveloped.

As shown in *Table 1*, the partial correlation for GRP and the economic complexity index is insignificant, while in the case of the partial correlation for ranks, there is a statistically significant relationship for these variables (hypothesis accepted at 5% level). This indicates the existence of a non-linear relationship between the index of economic complexity and GRP.

Let us take a closer look at the form of dependence of the logarithm of GRP on the ranks of economic complexity (*Fig. 4*).

Table 1.

Statistical estimates of partial correlations with economic complexity (for original variables and their ranks) for 2019. Results the family of hypothesis testing for equality of partial correlations to zero

	Partial correlation with economic complexity	H0: partial correlation with economic complexity is zero	Partial correlation with economic complexity for ranks	H0: partial correlation with economic complexity for ranks is zero
Fixed assets	0.16	1	0.20	1
Average annual number of employed persons	–0.30	0	–0.06	1
GRP	–0.14	1	–0.24	0
Researchers	0.03	1	–0.05	1
Extractive industry index	–0.53	0	–0.61	0
Manufacturing industry index	–0.12	1	–0.01	1

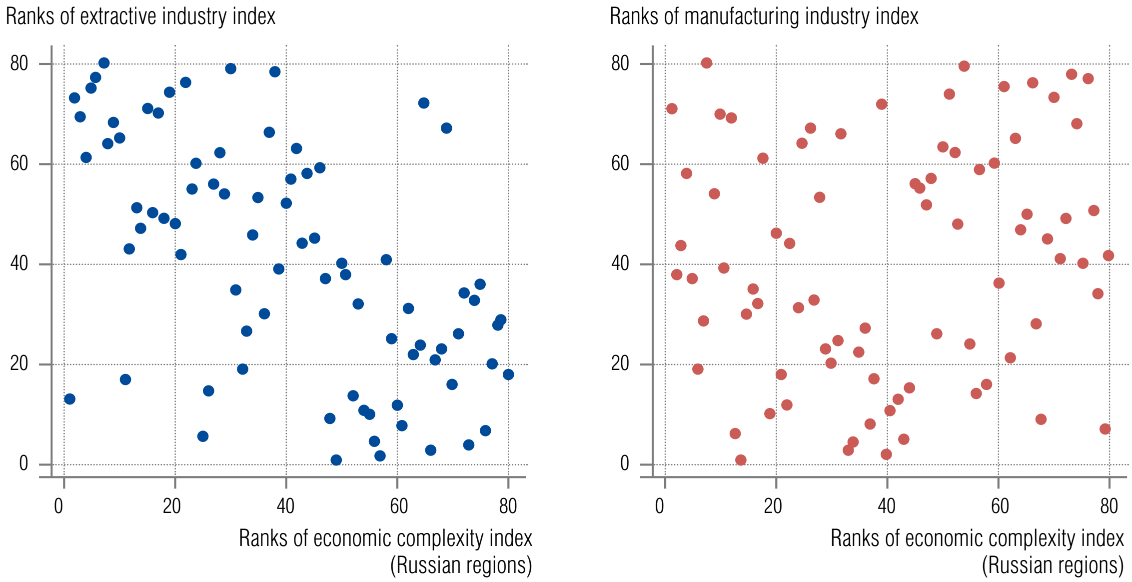


Fig. 3. Right: index of extractive industry (ranks) and economic complexity (ranks).
Left: index of manufacturing industry (ranks) and economic complexity (ranks).

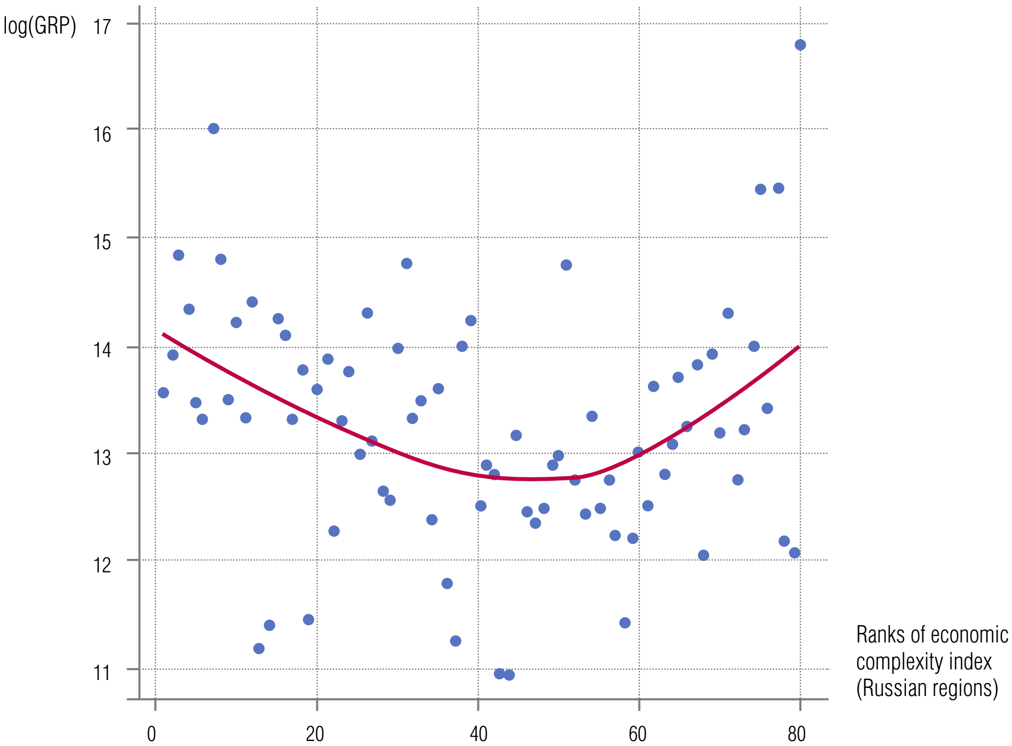


Fig. 4. GRP and ranks of economic complexity for the year 2019;
nonparametric Nadaraya–Watson regression (3).

Note that the U-shaped dependence in *Fig. 4*, obtained using nonparametric Nadaraya–Watson regression, suggests that there cannot be a simple linear relationship between economic complexity and GRP.

A U-shaped relationship between economic complexity and GRP implies that both very low and very high levels of economic complexity correspond to high GRP, while medium levels of economic complexity correspond to lower values of GRP. Thus, we can distinguish the following types of regional economies:

1. Low economic complexity, high GRP: regional economies tend to be rich in natural resources and their GRPs are heavily concentrated in extractive industries such as oil, gas or mining. Despite the low complexity of their economies (as they are mainly focused on one or a few sectors), these regions can have high GRPs due to the high market value of their resources and intensified mining.

2. High complexity, high GRP: regional economies with high economic complexity tend to have a wide range of developed and interconnected industries that include high-tech industries. These regions are typically characterized by high levels of industrialization, investment in human capital and technological innovation.

3. Medium level of complexity, lower GRP: regional economies that are in the process of transitioning to a more diversified and complex economy. There is a lack of developed capacity to efficiently produce more complex goods and services.

Thus, according to *Fig. 4*, we can distinguish two possible paths to higher GRP: (i) through natural resource extraction or (ii) through the development of a more sophisticated industrialized economy. Each pathway has its own advantages and challenges. For example, resource-rich regions may achieve high GRP quickly, but they may face instability due to fluctuations in commodity prices and may have difficulty diversifying their economies.

Due to the non-monotonicity of the correspondence between the logarithm of GRP and economic complexity, we take as a threshold for economic complexity the argument x_{opt}, at which the minimum of the constructed nonparametric Nadaraya–Watson regression (3) $g_{\tau_{opt}}(x)$ is reached (*Fig. 4*):

$$x_{opt} = \underset{x}{\operatorname{argmin}}\, g_{\tau_{opt}}(x) = 46.$$

The rank x_{opt} corresponds to an economic complexity value of 0.45.

Let's estimate the threshold impact of economic complexity on GRP:

$$cor(\text{GRP}, ECI \mid ECI \geqslant 0.45, X_{-(\text{GRP}, ECI)}) = 0.79,$$

$$cor(\text{GRP}, ECI \mid ECI < 0.45, X_{-(\text{GRP}, ECI)}) = -0.18,$$

where $X_{-(\text{GRP}, ECI)}$ all considered indicators of science and economy except GRP and ECI.

Thus, only at values of economic complexity exceeding 0.45, there is a direct relationship between GRP and the economic complexity index.

Based on the identified threshold direct relationship between economic complexity and GRP, the representation of the extended production function for GRP was summarized:

$$Y = c \cdot K^{\beta_1(S_1, S_2)} L^{\beta_2(S_1, S_2, T)} P^{\gamma} + \epsilon, \qquad (8)$$

where

$$\beta_1(S_1, S_2) = \frac{\mu_1 e^{(\mu_2 \cdot S_1 + \mu_3 \cdot S_2)}}{1 + \mu_1 e^{(\mu_2 \cdot S_1 + \mu_3 \cdot S_2)}}, \beta_2(S_1, S_2, T) =$$

$$= \frac{\lambda_1 e^{(\lambda_2 \cdot S_1 + \lambda_3 \cdot S_2 + \lambda_4 \cdot T^2)}}{1 + \lambda_1 e^{(\lambda_2 \cdot S_1 + \lambda_3 \cdot S_2 + \lambda_4 \cdot T^2)}};$$

$$T = \begin{cases} ECI, \text{ if } ECI \geq 0.45 \\ 0, \text{ otherwise}; \end{cases}$$

$c, \gamma, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \mu_1, \mu_2, \mu_3$ – constants;

Y – gross regional product in 2019;

K – fixed assets at the end of 2019;

L – average annual employment for 2019;

P – number of researchers for 2019;

ECI – economic complexity index calculated from data for 2019;

S_1 and S_2 – indices of extraction and manufacturing, respectively, calculated for 2019;

ϵ – errors of model (8).

Note that the expression found for GRP is estimated with greater accuracy, namely $R^2 = 0.982$, which is greater than in [16].

Endogeneity in model (8) occurs when the error ϵ is statistically dependent on one or more explanatory variables among K, L, P, S_1, S_2, T. Namely:

$$E(\epsilon \mid K, L, P, S_1, S_2, T) \neq 0.$$

It is well known that the presence of endogeneity leads to bias and invalidity of the LSE-estimators of the model parameters, leading to incorrect conclusions about the statistical significance of the relationships. To test the hypothesis about the absence of endogeneity is equivalent to the hypothesis testing for independence of each explanatory variable and errors in model (8). To test independence, we use the Hilbert–Schmidt independence criterion [25]. In contrast to the statistics in the Hausman test for independence of explanatory variables and the residuals of the model, which is assumed to be linear [26], the Hilbert–Schmidt test for independence allows for the presence of nonlinearity. A high value of the Hilbert–Schmidt independence criterion for a pair of variables indicates their dependence, while a low value corresponds to independence. Assuming that the null hypothesis is independence of the pair of variables under consideration, *Table 3* presents the results of the test.

As can be seen from the results of *Table 3*, the hypothesis that errors in model (8) are independent of explanatory variables is not rejected.

In order to make sure that the observed lack of relationship between errors and explanatory variables is not due to confounding variables, we test for conditional independence. For this purpose, we can also use the Hilbert–Schmidt independence criterion (*Table 4*).

Thus, according to the results presented in *Table 4*, the hypothesis of conditional independence of errors and explanatory variables in model (8) is also confirmed.

Table 2.

Parameter estimates of model (1) and their statistical significance

	Estimate	Sd. error	*t*-value	*p*-value	
C	6.77	0.42	4.53	0.00	***
μ_1	1.79	0.21	2.72	0.01	**
μ_2 (extractive industry index; fixed funds)	0.01	0.00	3.53	0.00	***
μ_3 (manufacturing industry index; fixed funds)	–0.02	0.01	–3.68	0.00	***
λ_1	0.33	0.26	–4.35	0.00	***
λ_2 (extractive industry index; employed)	–0.01	0.01	–1.96	0.05	*
λ_3 (manufacturing industry index; employed)	0.05	0.01	3.83	0.00	***
λ_4 (economic complexity)	3.34	1.16	2.89	0.01	**
γ (researchers)	0.05	0.02	2.81	0.01	**

Denotations: *** – p-value at less than 0.001 level, ** – p-value at less than 0.01 level, * – p-value at less than 0.05 level.

Table 3.

Testing the hypothesis of independence of the errors of model (8) and its explanatory variables

Pairs of variables	Hilbert–Schmidt independence criterion	*p*-value	Existence of independence
ϵ, T	0.0000033	0.13	independent
ϵ, K	0.0000364	0.8	independent
ϵ, L	0.0000366	0.79	independent
ϵ, P	0.0000257	0.97	independent
ϵ, S_1	0.000193	0.27	independent
ϵ, S_2	0.000237	0.23	independent

Table 4.

Testing hypotheses about conditional independence of errors of model (8) and its explanatory variables

Pairs of variables | condition	Hilbert–Schmidt independence criterion	*p*-value	Existence of independence
$(\epsilon, T \| T, L, P, S_1, S_2)$	0.00000154	0.11	independent
$(\epsilon, K \| T, L, P, S_1, S_2)$	0.0000503	0.98	independent
$(\epsilon, L \| T, K, P, S_1, S_2)$	0.000258	0.45	independent
$(\epsilon, P \| T, K, L, S_1, S_2)$	0.000149	0.89	independent
$(\epsilon, S_1 \| T, K, L, P, S_2)$	0.000368	0.77	independent
$(\epsilon, S_2 \| T, K, L, P, S_1)$	0.000657	0.18	independent

The presence of a statistically significant positive parameter λ_4 at truncated economic complexity suggests the possibility of the effect of "spillover" of innovations. Regions with a more complex production structure tend to have greater diversification, which creates opportunities for inter-sectoral diffusion of knowledge and technology, which in turn can lead to increased innovation and productivity growth. In addition, a region that produces a variety of products and has interconnected production processes has more opportunities to exploit economies of scale.

Figure 5 illustrates that increasing returns to scale are characteristic of regions with high elasticity of labor and low elasticity of capital.

As can be seen from *Fig. 5*, the growth of labor elasticity is accompanied by a decrease in the elasticity of capital and vice versa. This indicates a shift in the production function due to sectoral differences of regional economies.

Figure 6 shows that the presence of diminishing returns to scale is characteristic of regions with a high concentration of extractive industries in the structure of the regional economy. Declining returns to scale mean that a proportional increase in labor and capital leads to a less than proportional increase in output. This may be because extractive industries (e.g. mining, oil and gas) are often capital intensive and may face problems such as resource depletion, environmental regulations or high operating costs.

Figure 7 shows that increasing returns to scale are characteristic of regions with high manufacturing concentration and high economic complexity. Sufficiently large values of economic complexity, exceeding the threshold of 0.45, correspond to large returns to scale.

Since the economic complexity index characterizes the concentration of related sectors in the structure of an economy, as an economy becomes more complex, networking or relatedness facilitates the sharing of best practices and collaboration on innovation, and hence contributes to higher productivity. In a complex economy characterized by intricate inter-sectoral linkages and advanced production, the wealth of diverse knowledge and skills tends to be high. Sectoral relatedness allows this knowledge to be transferred between sectors, contributing to overall productivity. With high levels of sectoral relatedness, innovations and technological advances are more easily diffused across related sectors.

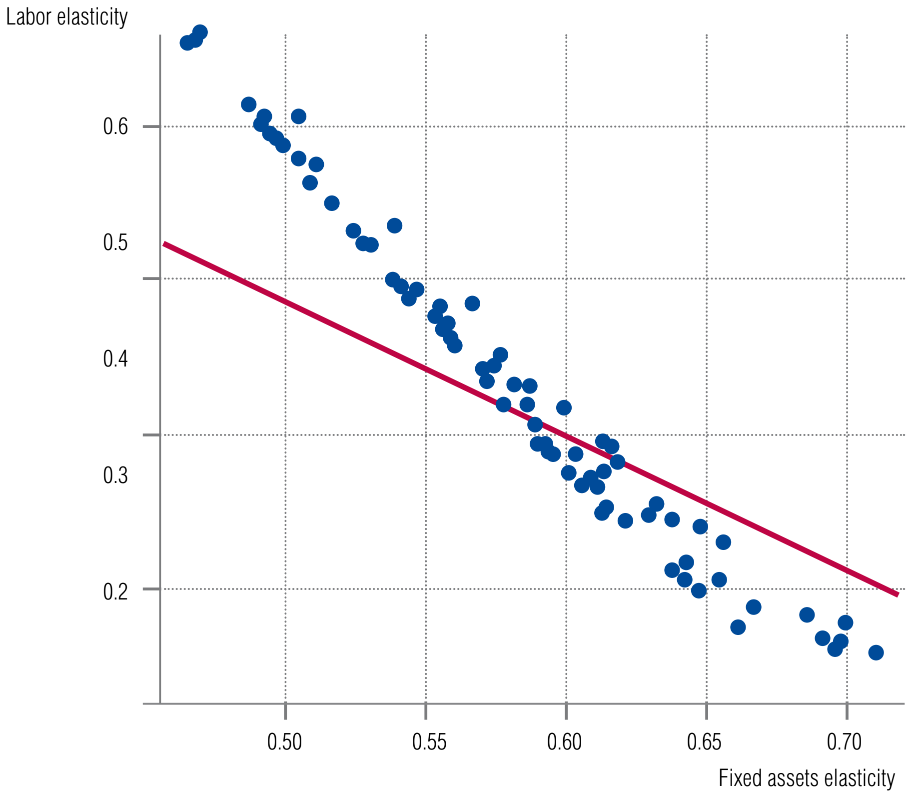


Fig. 5. For each region of Russia according to model (1):
$\beta_1(S_1, S_2)$ – elasticity of fixed assets, x-axis; $\beta_2(S_1, S_2, T_1,)$ – the elasticity of labor, y-axis.
Straight line: $x + y + \gamma = 1$.

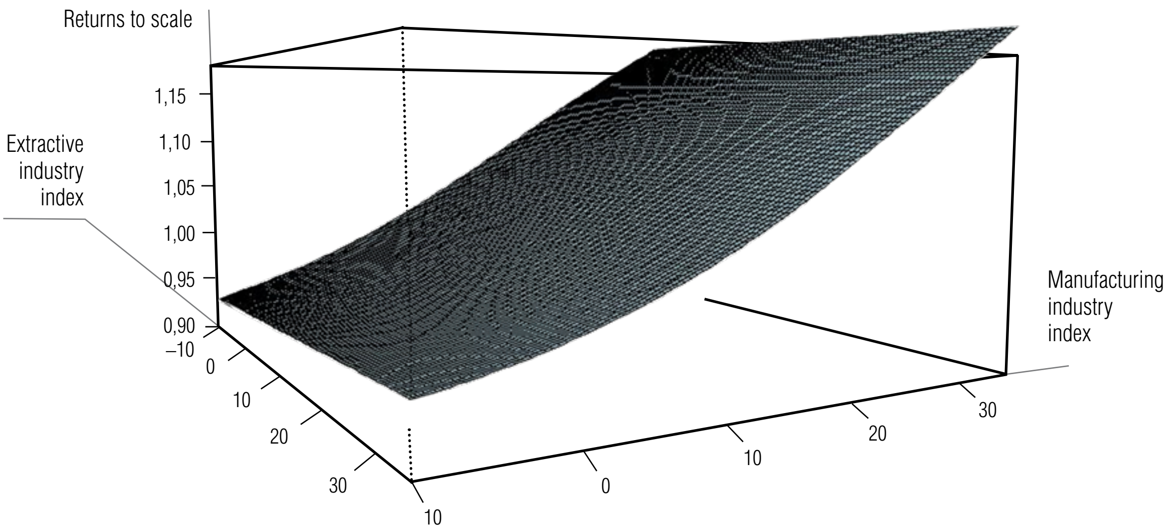


Fig. 6. Extractive industry index (*x*-axis); manufacturing industry index (*y*-axis); returns to scale (*z*-axis) calculated as $\beta_1(S_1, S_2) + \beta_2(S_1, S_2, T_1,) + \gamma$.

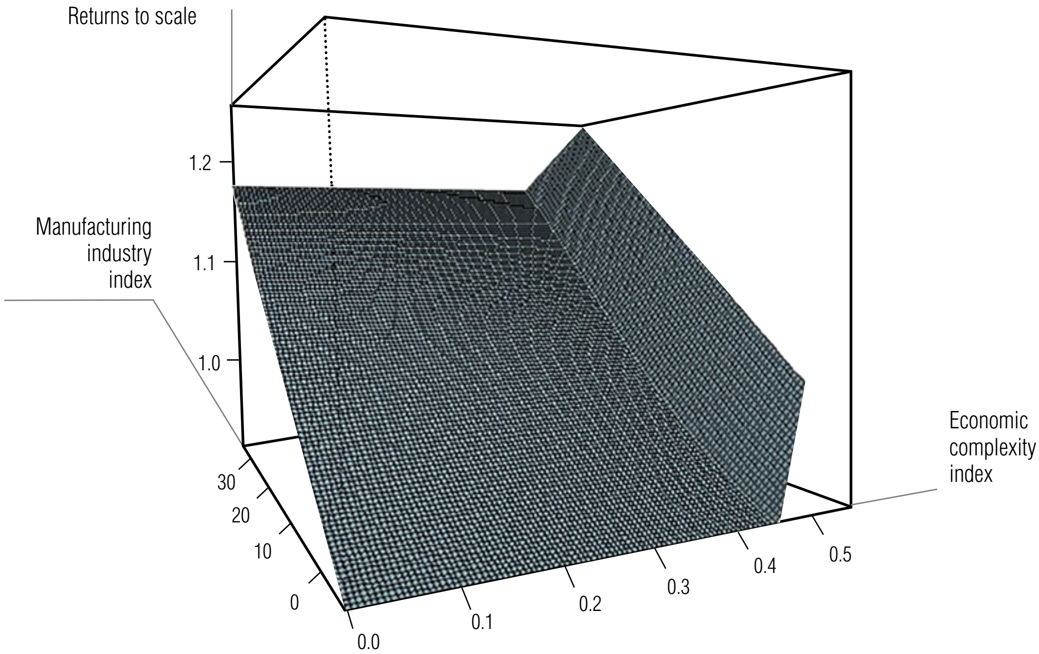


Fig. 7. Economic complexity index (*x*-axis); manufacturing index (*u*-axis); returns to scale (*z*-axis) calculated as $\beta_1(S_1, S_2) + \beta_2(S_1, S_2, T_1,) + \gamma$.

The formation of related sectors encourages co-development, where sectors do not grow in isolation, but through the joint development of technology, skills and knowledge. Such interconnected growth can further increase productivity through synergies between different sectors.

Thus, regions with more diverse and complex production structures with specialization in manufacturing are better positioned to benefit from economies of scale and respond to economic change. As noted earlier, regions with more complex economic structures tend to have more diversified economies, making them more adaptable to volatile economic conditions.

Conclusion

The most important results of the econometric study of the impact of economic complexity on the GRP of the Russian Federation regions, performed through the consistent use of three statistical methods (partial-correlation analysis of identifying direct relationships between variables, the Nadaraya–Watson method of nonparametric regression estimation, and the least squares method for nonlinear production functions) based on the statistical data for the year 2019, are as follows:

- There is no direct statistical relationship between the manufacturing industry index and economic complexity for the regions of the Russian Federation. This means that the emergence of new manufacturing sectors or the expansion of previously existing ones is not necessarily accompanied by an increase in economic complexity.
- The extractive industry index has a direct relationship with the economic complexity index. An increase in the extractive industry index corresponds to a decrease in the economic complexity index.
- Statistical evaluation of non-parametric Nadaraya–Watson regression showed that there is a non-linear relationship between GRP and the economic complexity index.
- Having ranked the regions by the level of economic complexity and excluding the influence of other variables in the sample, the rank number and the corresponding level of economic complexity were found, above which there is a direct relationship between GRP and economic complexity, and below which there is no such relationship.
- Statistical estimates of the parameters of the considered generalized production function show that the elasticity of fixed assets depends in a statistically significantly manner on the indices of sectoral specialization, while the elasticity of labor depends both on the indices of sectoral specialization and economic complexity. For values of economic complexity above a certain threshold, high economic complexity corresponds to higher labor elasticity. This indicates that regions with a more complex, diverse and interconnected production structure have higher productivity and, consequently, have more opportunities for efficient use of their labor resources.
- Increasing returns to scale are evident only in regions where manufacturing industries predominate and there is a sufficiently high level of economic sophistication. Regional economies with a high concentration of extractive industries are characterized by decreasing returns to scale, which potentially limits their growth.
- Manufacturing can provide more opportunities for productivity and value addition than extractive industries.

In general, the results of the study of the relationship between GDP and economic complexity emphasize the importance of taking into account economic complexity as an explanatory variable of the production function for regional GRP in its generalized form. Stimulating increases in economic complexity can be an effective way to promote economic growth and productivity, but this effect is only evident when the level of economic complexity is high enough. By increasing the diversity and economic complexity of their production structures, regions can increase productivity, competitiveness and economic stability, leading to higher levels of GRP and sustainable economic growth.

The importance of the composition of economic sectors and the balance between labor and capital in shaping output and growth should also be emphasized.

Regions concentrated in sectors with high labor elasticities are characterized by increasing returns to scale and hence potentially higher economic growth. Conversely, regions with a high concentration of extractive industries may experience declining returns to scale, potentially limiting their growth. This underscores the importance of policies aimed at increasing productivity and diversification away from extractive industries for sustainable economic growth.

The methodology presented in this paper for quantifying the impact of economic complexity on gross regional product (GRP) can be useful in the decision-making process of locating new production facilities, distribution centers or branches of enterprises. Understanding the impact of economic complexity on GRP can help identify economically stable and sufficiently diversified regions with more favorable business conditions. However, regions with a complex economic structure are also characterized by a higher potential level of competition.

Higher GRP usually correlates with higher purchasing power of consumers. Therefore, the results presented may help businesses to identify potentially lucrative regional markets for their products or services. However, it is important to note that, while useful, this methodology is one tool and should be used in conjunction with other data sources and market research to make comprehensive decisions. ■

References

1. Hirschman A.O. (1958) *The strategy of economic development*. New Haven: Yale Univ. Press.
2. Rosenstein-Rodan P.N. (1943) Problems of industrialization of eastern and southeastern Europe. *The Economic Journal*, vol. 53(210/211), pp. 202–211.
3. Teece D., Rumelt R., Dosi G., Winter S. (1994) Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior & Organization*, vol. 23(1), pp. 1–30.
4. Bernard A.B., Jones C.I. (1996) *Productivity and convergence among U.S. States*. National Bureau of Economic Research.
5. Imbs J., Wacziarg R. (2003) Stages of diversification. *The American Economic Review*, vol. 93(1), pp. 63–86.
6. Boschma R., Iammarino S. (2009) Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, vol. 85(3), pp. 289–311.
7. Frenken K., Oort F.V., Verburg T. (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies*, vol. 41(5), pp. 685–697.
8. Frenken K., Saviotti P. (2008) Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, vol. 18(2), pp. 201–218.
9. Hausmann R., Hidalgo C.A., Bustos S., Coscia M., Chung S., Jimenez J., Simoes A. (2007) The building blocks of economic complexity. *PLoS One*, vol. 2(1), e268.
10. Hausmann R., Hidalgo C.A. (2017) *Atlas of economic complexity: Mapping paths to prosperity*. MIT Press.
11. Hidalgo C.A., Hausmann R. (2009) The building blocks of economic complexity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367(1897), pp. 1817–1825.
12. Hausmann R., Hidalgo C.A., Bustos S., Coscia M., Simoes A., Yildirim M.A., Hwang J. (2014) *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.
13. Ravallion M. (2004) Competing concepts of inequality in the globalization debate. *Brookings Trade Forum Globalization, Poverty, and Inequality*, pp. 1–38. Available at: https://www.jstor.org/stable/25063189 (accessed 14 February 2023).

14. Hartmann D., Guevara M.R., Jara-Figueroa C., Aristarán M., Hidalgo C. (2017) Linking economic complexity, institutions, and income inequality. *World Development*, vol. 93, pp. 75–93.
15. Hidalgo C.A., Klinger B., Barabási A.L., Hausmann R. (2007) The product space conditions the development of nations. *Science*, vol. 317(5837), pp. 482–487.
16. Gavrilets Y.N., Kudrov A.V., Tarakanova I.V. (2018) Analysis of the internal structure of economic growth potential. *Bulletin of CEMI RAS*, vol. 1, no. 1 (in Russian).
17. Gavrilets Y., Kudrov A., Tarakanova I. (2022) Statistical analysis and modeling of regional economy and science relationship. *Ekonomika i matematicheskie metody*, vol. 58(4), pp. 56–70. https://doi.org/10.31857/S042473880023019-9 (in Russian).
18. *Regions of Russia. Socio-economic indicators, 2017–2019*. Moscow: Rosstat (in Russian).
19. *2019–2020 tax base and accrual structure reports for taxes and levies*. Available at: https://www.nalog.gov.ru/rn77/related_activities/statistics_and_analytics/forms/ (accessed 14 February 2023) (in Russian).
20. Afanasiev M., Kudrov A. (2021) Economic complexity and embedding of regional economies' structures. *Ekonomika i matematicheskie metody*, vol. 57(3), pp. 67–78. https://doi.org/10.31857/S042473880016410-0 (in Russian).
21. Bierens H.J. (1994) *Topics in advanced econometrics estimation, testing, and specification of cross-section and time series models*. Cambridge University Press.
22. Bishop C.M. (2006) *Pattern recognition and machine learning*. Springer.
23. Ratkowsky D. (1993) Principles of nonlinear regression modeling. *Journal of Industrial Microbiology*, vol. 12, pp. 195–199.
24. Seber G.A.F., Wild C.J. (2003) *Nonlinear Regression*. New York: Wiley–Interscience.
25. Gretton A., Fukumizu K., Harchaoui Z., Sriperumbudur B.K. (2009) A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, pp. 1067–1074.
26. Hausman J.A. (1978) Specification tests in econometrics. *Econometrica*, vol. 46(6), pp. 1251–1271.

About the author

Alexander V. Kudrov

Cand. Sci. (Phys.-Math.);

Leading Researcher, Central Economic and Mathematical Institute of the Russian Academy of Sciences, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: kovlal@inbox.ru

ORCID: 0000-0003-2495-5496

DOI: 10.17323/2587-814X.2023.4.41.56

Mathematical model of the formation of supply chains of raw materials from a commodity exchange under conditions of uncertainty*

Rodion S. Rogulin
E-mail: rafassiaofusa@mail.ru

Vladivostok State University
Address: 41, Gogolya Sreet, Vladivostok 690014, Russia

Far Eastern Federal University
Address: 10, Ajax Village, Russky Island, Primorskiy Territory 690922, Russia

Abstract

The formation of raw material supply chains is very closely related to production problems at a timber processing plant. Since the beginning of the second industrial revolution, one urgent question has been the formation of supply chains for raw materials and the optimal calculation of production volumes for each individual day. This article examines a forestry enterprise that does not have its own sources of wood, which daily solves the problem of ensuring the supply of raw materials and optimal production load. A commodity exchange is considered as a source of raw materials where lots randomly appear every day in different raw material regions. In the scientific literature, there are many approaches to calculating

* The article is published with the support of the HSE University Partnership Programme

the optimal profit value over the entire planning horizon, but they do not consider many features that are important for a timber processing enterprise. This paper presents a mathematical model which is a mechanism for making daily decisions over the entire planning horizon and differs in that it allows one to take into account the share of useful volume and the delivery time of raw materials under conditions of uncertainty. The result of the model is the optimal profit trajectory, considering the volume of raw materials, the delivery time of lots, the volume of profit and the production volume of goods. The model was tested on data from the Russian Commodity and Raw Materials Exchange and one of the Primorsky Territory enterprises. Analysis of the results showed that there are difficulties in planning supply chains and production volumes. An assessment of the optimality of raw material regions was carried out. The advantages and disadvantages of the mathematical model are formulated.

Keywords: production optimization, transport problem, forest industry, commodity exchange, supply chains, product release

Citation: Rogulin R.S. (2023) Mathematical model of the formation of supply chains of raw materials from a commodity exchange under conditions of uncertainty. *Business Informatics*, vol. 17, no. 4, pp. 41–56.
DOI: 10.17323/2587-814X.2023.4.41.56

Introduction

The wood supply chain plays a vital role in the global economy, providing essential raw materials to various industries such as construction [1–4], furniture [5–7] and paper [8]. Supply chains are complex and dynamic; their formation depends on the efficiency of forestry management, processes of harvesting, wood processing, distribution and consumption of wood products. Effective supply chain management (SC, SCM) of wood is critical to ensure sustainability of production and conservation of resources while promoting economic growth.

In recent years, there has been increased interest in wood supply chain modeling to optimize and improve production sustainability [9–11]. Modern SC formation models are developed to provide management with valuable information about supply chain performance and can be used to improve the sustainability and efficiency of management strategies [12, 13]. There is a significant amount of research, the scope of which lies in various areas of forestry production: forest management and the processes of logging, processing, distribution and consumption.

Modeling wood supply chain management processes is critical to efficiently use resources, reduce environmental impact and improve economic performance. As technology advances, data analysis techniques such as machine learning and stochastic optimization are being used to develop models to provide more accurate predictions and effective decision-making tools.

1. Literature review

The wood supply chain is a complex network of processes including forest management, harvesting, transportation, processing and distribution. Effective wood supply chain management is critical to optimizing operations and achieving the sustainability goals of any forest products business. Mathematical modeling is a powerful tool to support decision-making processes and optimize various aspects of wood supply chain management. Let's consider the experience of modeling the problem of supply chain management.

1.1. Timber harvesting optimization models

Optimizing timber harvesting processes is essential for sustainable forest management and maximizing economic returns. Mathematical models such as Mixed-Integer Linear Programming (MILP) models are widely used to determine the optimal timing and spatial distribution of logging. Such models consider various factors, including wood growth, market demand, operating costs and environmental constraints, to support decision-making processes. For example, in [9], the authors propose a MILP model that optimizes timber harvesting and road construction operations in several logging areas, considering economic, environmental and transportation factors. The model aims to maximize the net present value of timber revenues while minimizing logging costs and environmental impacts.

Efficient logistics and transportation are critical components of the wood supply chain, ensuring timely delivery of wood products and minimizing transportation costs. Mathematical modeling methods are used to optimize transport routes, vehicle planning and inventory management. The authors of the study [10] developed an MILP model for multi-day truck routing. The model was applied to a Brazilian logging company that operated one factory and several logging sites equipped with cranes. The model considers the Less Than Truckload (LTL) problem, allowing for repeated movement of trucks between several forest apiaries. The objective function aims to minimize transportation costs, the number of trucks, the number of trips and overtime costs. The model was applied to a case with 5 available harvesters and 48 trucks. Methods that directly use optimization models are sensitive to the dimension of the problem in terms of the number of variables and constraints. In addition, such models usually require assumptions, for example about full trucks, as in [10], to be able to formulate a mathematical model of the problem being solved.

1.2. Inventory and supply chain management models

There are many models in the literature dedicated to the development and application of various approaches to solving a variety of problems in the field of inventory and supply chain management. For example, intuitionistic fuzzy sets (hereinafter IFS) were used [14, 15]. Measures of possibility, necessity and reliability are used as a new approach to solving intuitionistic fuzzy optimization problems [16, 17]. They are also applied in manufacturing, stock-out inventory models to obtain Pareto optimal solutions [18]. Multi-objective IFS optimization has been used, for example, in studies [19–21].

In addition, dynamic programming methods were also used to optimize multi-echelon supply chain problems. Thus, a neurodynamic programming model was developed in [22] to solve the two-stage problem of inventory optimization under conditions of demand uncertainty. Testing of the model in practice showed a reduction in enterprise costs by 10%. The authors of the study [23] formulated the supplier-managed inventory routing problem as a Markov decision process and applied the approximate dynamic programming method to solve it. The authors of [24] developed an approximate dynamic programming (ADP) model based on Lagrangian relaxation for inventory management of a network with one product and several sites. The authors of [25] use ADP methods and apply stochastic approximation to calculate optimal underlying inventory levels given news provider problems over the horizon of multiple periods of backlogs and lost sales. The authors of the study [26] apply ADP methods to solve the problem of inventory management at several enterprises and with a given number of products, considering the variability of some processes.

Reinforcement learning has also been applied to the Inventory Management Problem (IMP) [27]. Thus, the team of authors in [28] use Q-learning for a four-stage IMP with a 12-week cycle and non-stationary demand. In [29], the authors trained the Deep Q-Network neural network architecture to achieve near-optimal results in the Beer Game, a classic example of a multi-layer IMP. The authors of [30] uses Q-learning methods and the SARSA model for optimal replenishment of perishable goods.

1.3. Conclusions and formulation of the research problem

Modeling has become an important tool for analyzing and improving supply chains in various manufacturing areas. The complexity of the supply chain, along with increasing pressure to minimize environmental impact, implement sustainable practices and consider social and operational considerations, has led to the development of a wide range of models. In the wood supply chain, modeling can optimize the flow of raw materials from their origin to their destination by minimizing costs, reducing environmental impact and increasing efficiency.

As a review of the literature showed, there are many works devoted to the subject of SCM, however, many models and approaches are not applicable in practice when managing a forestry enterprise in matters of forming supply chains in conjunction with determining production volumes.

It is worth noting that the existing models are affected by the failure, firstly, to take into account the coefficient of the useful volume of raw materials that will reach the warehouse. This must first be separated from rot, and then processed into dust. Then one uses a press to produce OSB boards. Secondly, there are the tools of daily decision-making based on the supply of lots of raw materials on the exchange that day. For example, work [12] is devoted to a similar problem, but does not take into account the considered feature with the useful volume of raw materials that will reach the warehouse. In addition, there is the absence of a qualitative forecast of the situation on the commodity exchange (when and in what volume lots will appear for sale on the stock exchange) and it is not at all applicable in practice. Work [11] is devoted to the same problem, but with an emphasis on pricing of final goods. Here, as in [12], there is no consideration of the coefficient of the useful volume of raw materials that can reach the warehouse. In addition, the second negative side of the model is repeated: there is no possibility to make a daily decision based on the current offer on the stock exchange. However, it is worth noting that the problem of considering the coefficient of useful volume of raw materials is not new and was considered in [13]. This article is devoted to assessing the optimal value of profit over the entire planning horizon of the forestry production cycle, namely: SCM and calculation of production volumes, where the flow of raw materials is carried out from the commodity exchange. However, there is still no clear description of how to use the model so that the enterprise can make decisions related both to the problems of purchasing raw materials from the exchange and to calculating production volumes so that the profit value at the end of the planning horizon is maximum and at the same time extremely close to optimal.

Thus, the problem of managing an enterprise in the timber industry remains unresolved, when the enterprise does not have its own sources of raw materials for production, and when every day a decision has to be taken on the formation of the flow of raw materials from the commodity exchange and at the same time on the volume of finished products over the entire planning horizon, maximizing the value of the total profits, in conditions of uncertainty in the supply of lots on the stock exchange and taking into account the coefficient of useful volume of raw materials, technology of production of goods and the specifics of logistics of raw materials to the warehouse. This work is devoted to the development of a model that solves the current problem.

2. Research purpose, objectives and hypothesis

Considering the processes of enterprise functioning, the most important for forestry production are the formation of supply chains for raw materials and the volume of production of goods.

Considering the sources of raw materials entering the stock exchange, the exchange enters into agreements with tenants of forest plots from various regions on the use of the trading platform. After completing an exchange transaction between the raw material processing enterprise (timber industry complex) – the customer and the tenant of the forest plots (logger) – the seller, the volume of raw materials stated in the contract is sent to the customer. It is only possible to buy a lot on the exchange in its entirety [11–13, 31–33].

The purpose of this study is to develop a mathematical model that allows the formation of supply chains of raw materials from the commodity exchange in conditions of supply uncertainty and differs from already known models in that, firstly, the decision-making process is carried out daily, and, secondly, it allows you to take into account the likely time of delivery[1] of raw materials to the warehouse and changes in the working volume of raw materials which depends on external factors (temperature, insects, etc.).

Research objectives:

1. Construction of an economic and mathematical model.
2. Designing an algorithm for finding a solution for the developed model.
3. Analysis of model testing results.

It is well known that such a problem can be solved optimally when all the values that were played (lots, travel time) are already known. However, there is no understanding whether it is possible to make decisions every day on the formation of supply chains of raw materials and on production volumes so that the profit value is as close to the optimal value as possible. We assume that a model which allows you to solve the problem as close as possible to the optimal one exists, where decisions are made every day, having only the data of the current day and many assumptions about what the situation will be "tomorrow."

3. Mathematical model[2]

Any production, including the timber industry, is unable to function without the existence of a source of raw materials. To ensure the supply of the required volume of raw materials, it is necessary to identify wood suppliers. To do this, we will use the services of the St. Petersburg International Commodity and Raw Materials Exchange (SPbIMRME)[3]. Every day the stock exchange publishes data on how many transactions (orders) were made, at what price and what volume of raw materials was sold. In addition, the exchange provides services for the delivery of raw materials to the consumer, which is also included in the price of the goods. The exchange represents many regions from which raw materials could potentially come [11–13, 33]. We will specifically change or expand the input data for solving the problem to make it more difficult for the model to find a solution.

After a sufficient volume of raw materials has arrived at the production warehouse, the enterprise must decide on the optimal vector to produce the final product, focusing on the maximum possible production volume [11–13, 33].

Let's consider the scheme for purchasing raw materials and calculating production volumes. It is known that profit maximization at an enterprise is achieved if and only if the necessary quantities are calculated in a single model. That is to say, the model considers both the volume of production and the flow of raw materials. Since an enterprise usually does not know what will happen on the market (exchange) tomorrow, it plans only today based on the estimated situation at the enterprise "tomorrow."

This raises the question of which planning period $\tilde{T} \geqslant 1$ to choose. In this work we will set it to one value to solve the entire problem.

We introduce some assumptions. Let, firstly, the enterprise know for a certain value of E periods what lots were drawn on the stock exchange and the situation with workload on the railways (railroads). Secondly, we also assume that the situation on the raw

[1] The purchase and sale agreement specifies the methods and price of timber delivery. Delivery can be carried out by the enterprise, but further we will consider the delivery of raw materials by the supplier.

[2] The software implementation of the developed model and exchange sales statistics can be found at the link https://drive.google.com/drive/folders/1THzU7BHjGgpUgZiXQbvJNaIA14biWO1t?usp=sharing

[3] https://spimex.com/

materials market does not change much over the years. Then the enterprise, based on these data, with information at one's enterprise for the same periods, can build a model to find the optimal solution. Then there is an opportunity to get many optimal trajectories of profit values and volumes of raw materials in the warehouse every day over the entire planning horizon. In this work, we focus on considering the trajectory of raw material inventories in a warehouse.

In this case, it is possible to build a regression that would reflect the average expected value in aggregate for all types of raw materials in warehouse $\tilde{b}_m$ on each individual day m over the entire planning horizon of M days, depending on which lots are available today.

Since the solution has to be sought on a certain interval $[m, \min(M, m+\tilde{T})]$, and the value of the expected total volume of raw materials is known only for the current day m $\tilde{b}_m$, the question arises about what volume of raw materials in the warehouse we expect in the next days $[m+1, \min(M, m+\tilde{T})]$, where $m+1 \neq M$.

Returning to already found optimal trajectories of the total volume of raw materials in the warehouse in the previous E periods, from here you can calculate the value

$$\tau(m)=\begin{cases}\dfrac{\sum_{e=1}^{E}\dfrac{\sum_l b_{lm}(e)}{\sum_l b_{l(m-1)}(e)}}{E}, m<M,\\ 0,\ m=M\end{cases}$$

where $\sum_l b_{lm}(e)$ – this is the total volume of raw materials for all their types l on day m for each individual input data sample $e \in E$. Then we will search for a solution to the current problem that would take into account the forecast value $\tilde{b}_m$ on the current day m and some correction for the entire planning period forward $\min(M-m, \tilde{T})$

$$\prod_{\underline{t}=1}^{t}\tau(m+\underline{t}),\ t=0: \min(M-m, m+\tilde{T}).$$

Let's introduce the following set of parameters and variables.

Options:

p_{km} – price for product type k on day m;

c_{ilrm} – price of lot i with type of raw material l from region r, appearing on the exchange on day m;

A_{lk} – rate of consumption of raw materials of type l for the production of a unit of goods of type k;

$\gamma_{\tilde{m}m}$ – spoilage rate of raw materials purchased on day $\tilde{m}$ to day m $(m \geqslant \tilde{m})$;

V_{ilrm} – volume of raw materials in lot i with raw material type l from region r, appearing on the exchange on day m;

H_{nk} – maximum production volume of goods of type k on day m;

$\underline{b}$ – emergency level of raw material reserves;

$\overline{b}$ – maximum warehouse capacity;

B_0 – initial budget level;

FC – fixed costs;

M – planning horizon;

L_r – distance from warehouse to region r;

S_m – distance traveled by the application on day m;

π_m – profit value at the moment day m;

$\varepsilon^{(3)}$ – noise (random variable) component of the working volume of raw materials that reached the warehouse;

left and *right* – the minimum and maximum value of a random variable distributed according to a uniform law;

$LN(a_m, \delta_m)$ – log normal distribution of a random variable with parameters (a_m, δ_m) respectively;

$\tilde{T}$ – the period for which the enterprise solves the task $F_m^{(1,2)}$ (days);

E – number of different sets of input parameters $\{V_{ilrm}(e), c_{ilrm}(e), T_{r\tilde{m}}(e)\}$.

Let's consider the notation with and without a tilde above the parameter. We will assume that the value with a tilde above the variable is the value that the enterprise evaluates, and without the tilde is its real value. For example,

$T_{r\tilde{m}}$ – time during which a lot purchased on day $\tilde{m}$ from region r will reach the warehouse;

$\tilde{T}_{r\tilde{m}}$ – time during which, according to the enterprise's estimates, a lot purchased on day $\tilde{m}$ lot from region r will reach the warehouse.

Variables:

x_{km} – volume of production of goods of type k on day m;

λ_{ilrm} – fact of purchase of lot i with type of raw material l from region r, which appeared on the exchange on day m;

b_{lm} – level of stock of raw materials of type l in the warehouse on day m;

$b_{lm}(e)$ – the value of the stock of raw materials of type l on day m, which was found when solving the problem on data e.

Task $F_m^{(1,2)}$ takes the form (1–18). Objective function (1) is aimed at maximizing profit values on each day $m + t - 1$:

$$\sum_{t=1}^{\tilde{T}}\left(\sum_k p_{k(m+t-1)}x_{k(m+t-1)} - \sum_{j=1}^{2}\sum_l N^{(j)}\varepsilon_{l(m+t-1)}^{(j)}\right) - \sum_{i,l,r} c_{imrl}\lambda_{imrl} \to \max. \quad (1)$$

Relation (2) specifies the relationship between the volumes of raw materials in the warehouse, the volume of raw materials spent in production and the volumes of raw materials that reached the warehouse:

$$b_{l(m+t-1)} - b_{l(m+t-2)} + \sum_k A_{lk}x_{k(m+t-1)} - \tilde{\gamma}_{\tilde{m}(m+t-1)}\sum_{i,r}V_{i\tilde{m}rl}\lambda_{i\tilde{m}rl} + \sum_{j=1}^{2}(-1)^j\varepsilon_{l(m+t-1)}^{(j)} = 0, \quad (2)$$

where

$t = 1: \tilde{T}$;

$\tilde{T} = \text{const}$;

$\tilde{T} \geq \max(\tilde{T}_{r\tilde{m}})$;

$N^{(j)} >> 1$;

$j = 1: 2$;

$b_{l(m-1)} = \text{const}$;

$\tilde{T} = \min(\tilde{T}, M - m + 1)$;

$\tilde{m} + \tilde{T}_{r\tilde{m}} = m + t - 1$.

Constraints (3–4) set possible limits for the values of variables responsible for the volume of production of each type of product on each day (3) and when the fact of purchasing raw materials on the exchange is a variable, and in which cases it is considered a constant (4). The $\lambda_{i\tilde{m}rl}$ variable responsible for the fact of making a decision to purchase lot i on day $\tilde{m}$ from region r with raw material type l is a constant if and only if $\tilde{m} \leq m$, where $\tilde{m}$ is the date of appearance of the lot in question on the exchange, and m is the day of acceptance solutions, otherwise is a variable:

$$x_{km} \in N, \quad (3)$$

$$\lambda_{i\tilde{m}rl} = \begin{cases} \text{const}, \tilde{m} \leq m \\ \{0; 1\}, \begin{cases} \tilde{m} = m \\ \tilde{T}_{r\tilde{m}} = t - 1. \end{cases} \end{cases} \quad (4)$$

Constraint (5) is aimed at ensuring that the search for a solution is based on the required total volume of raw materials in the warehouse. At the end of the planning horizon, we will assume that the enterprise stops its production, so it does not require raw materials in the warehouse:

$$\sum_l b_{l(m+t-1)} = \begin{cases} \min\left(\bar{b}, \prod_{\underline{t}=1}^{t}\tau(m+\underline{t})\cdot\tilde{b}_m\left(\{V_{imrl}\}_{i,r}, \{c_{imrl}\}_{i,r}, m\right)\right), \\ t \in (1, \tilde{T}) \\ 0, t = M - m + 1, \end{cases} \quad (5)$$

where

$$\sum_l b_{lm} = \tilde{b}_m\left(\{V_{imrl}\}_{i,r}, \{c_{imrl}\}_{i,r}, m\right);$$

$$\tau(m) = \begin{cases} \dfrac{\sum_{e=1}^{E}\dfrac{\sum_l b_{lm}(e)}{\sum_l b_{l(m-1)}(e)}}{E}, m < M \\ 0, m = M. \end{cases}$$

Inequality (6) states that there cannot be a situation where the volume of raw materials in the warehouse drops below zero, taking into account of penalty variables $\varepsilon_{l(m+t-1)}^{(j)}$:

$$b_l(m+t-1)+\sum_{j=1}^{2}(-1)^j \varepsilon_{l(m+t-1)}^{(j)} \geq 0. \quad (6)$$

Inequality (7) reflects the current level of the enterprise budget on each individual day $m + \underline{t} - 1$ during whole planning horizon $m + t - 1$:

$$\pi_{m-1}+\sum_{\underline{t}=1}^{t}\left(\sum_k p_{k(m+\underline{t}-1)} x_{k(m+\underline{t}-1)}\right)- -\sum_{i,l,r} c_{imrl}\lambda_{imrl} \geq FC \cdot t, \quad (7)$$

where $\pi_0 = B_0$.

Inequality (8) is intended to prevent a situation where the volume of raw materials in the warehouse exceeds the maximum capacity:

$$\sum_l b_{l(m+t)} \leq \bar{b}. \quad (8)$$

Formulas (9–10) is a system in which the drawn values reflect the maximum travel time of the lot:

$$\tilde{T}_{r\tilde{m}} = m^*: \begin{cases} \left| L_r - \sum_{\underline{m}=\tilde{m}}^{m^*} S_{\underline{m}} \right| \to \min \\ L_r - \sum_{\underline{m}=\tilde{m}}^{m^*} S_{\underline{m}} \leq 0 \end{cases} \quad (9)$$

$$S_{(m+t-1)} \sim LN\left(\tilde{a}_{(m+t-1)}, \delta_{(m+t-1)}\right). \quad (10)$$

Formulas (11–12) specify an estimate of the coefficient of the useful volume of raw materials that will reach the warehouse. Formula (11) uses the function arctg(x) to reflect the useful volume of raw materials that has reached the warehouse. Since the function y = arctg(x) can take values in the interval [0; $\pi/2$], $x \in [0; \infty]$, it is necessary to make changes to the relationship between the value of the coefficient[4] of raw materials reaching the warehouse and how much time has passed since the moment the lot appears on the exchange[5] in such a way that the condition is met: $0 \leq \text{arctg}(x) \leq 1$, $x \in [0; \infty]$, $\beta = \text{const}$, $\tilde{\beta} = \text{const}$:

$$\tilde{\gamma}_{\tilde{m}(m+t-1)} = \min\left(1; \max\left[0; 1-\frac{2}{\pi}\text{arctg}\left(\tilde{\beta}\left(\left(m+t-1\right)- -\tilde{m}\right)\right)+\tilde{\varepsilon}^{(3)}\right]\right), \quad (11)$$

$$\tilde{\varepsilon}^{(3)} \sim U(\widetilde{left}), (\widetilde{right}). \quad (12)$$

Formulas (13–14) limit the values of some variables:

$$0 \leq x_{k(m+t-1)} \leq H_{k(m+t-1)}, \quad (13)$$

$$\varepsilon_{l(m+t-1)}^{(j)} \geq 0, j = 1{:}2. \quad (14)$$

After the solution, the following key parameters are calculated (15–18):

$$\pi_m = \pi_{m-1} + \sum_k p_{km} x_{km} - \sum_{i,l,r} c_{imrl}\lambda_{imrl} - FC, \quad (15)$$

$$\gamma_{\tilde{m}(m+t-1)} = \min\left(1; \max\left[0; 1-\frac{2}{\pi}\text{arctg}\left(\beta\left(\left(m+t-1\right)- -\tilde{m}\right)\right)+\tilde{\varepsilon}^{(3)}\right]\right), \quad (16)$$

$$b_{lm} = b_{lm-1} - \sum_k A_{lk} x_{km} + \gamma_{\tilde{m}m}\left(\sum_{i,r} V_{i\tilde{m}rl}\lambda_{i\tilde{m}rl}\right), \quad (17)$$

where $\tilde{m} = m - \tilde{T}_{r\tilde{m}}$;

$$m = m + 1. \quad (18)$$

Recalculation of the main indicators occurs after the solution has been found for day m (15–17). In equality (18), the equal sign is used as an assignment operator to the value to the left of the equal sign to the value to the right.

4. Model calibration

To test the model, one of the leading forestry enterprises of the Primorsky Territory was chosen. Enterprises from four regions are most represented on the

[4] We will assume that this coefficient is in the range from 0 to 1.

[5] Let us introduce the assumption that the raw materials indicated in the lot were produced on the same day on which the lot was put up for auction.

exchange: Irkutsk Region ($r = 1$), Perm Region ($r = 2$), Republic of Buryatia ($r = 3$), Moscow Region $r = 4$). The planning horizon lies between February 1, 2022 and mid-May 2022.

Knowing the coordinates of enterprises, it would be possible to conduct a dialogue with them directly, bypassing the stock exchange. However, the exchange hides the real coordinates, so all transactions are carried out through the exchange, both from the buyer and from the supplier [11].

An array of the following data was collected from the official website of the exchange and from the enterprise for the specified period: prices of proposed applications c_{irm}, volumes of applications v_{ilrm}, selling prices of final goods p_{k1}, number of applications for each type of raw material. Since the exchange website runs software that does not allow data to be collected automatically, we can conclude that the data is not allowed to be used in large quantities, so all of the above values will be noisy and slightly changed.

The main initial data[6] characterizing the enterprise are presented in *Tables 1* and *2*.

For calculations, we will use the high-level programming language Matlab and the function from the intlinprog[7] extension package to find solutions to linear optimization problems.

Evaluating[8] function $\tilde{b}_m\left(\{V_{imrl}\}_{i\ r}, \{c_{imrl}\}_{i\ r}, m\right)$ with input parameters $\left\{\{V_{imrl}\}_{i,r}, \{c_{imrl}\}_{i,r}, m\right\}$, neural networks[9] (NN) with the following properties[10] were used: 10 hidden layers with activation function tg(x), 1 output layer (ReLU), learning algorithm – Levenberg–Marquardt. Coefficient of determination (Euclidean Metric) $R^2 = 0.78$.

Table 1.

Basic initial parameters of the enterprise

Parameters, units of measurement	Values
p_{km}, rub	$\forall m$: (1; 1,5; 1,6; 1,7) · 10^4
$N^{(j)}$,conventional units	$15p_{km}$
$\tilde{T}$, days	22
M, days	100
$\bar{b}$, m^3	7000
FC, rub	100 000
B_0, rub	3 000 000
L_r, km	(3740, 7560, 3250, 9000)
H_{nk}, units	$\forall m, k$: $H_{km} = 4$

Sources: enterprise, author.

Table 2.

Cost of raw materials per unit of production

		Item number (*k*)			
		1	2	3	4
Raw material type number (*l*)	1	2	3	4	3
	2	1	3	3	5

Sources: enterprise.

[6] The author does not have the right to publish some of the company's data due to his signing an agreement "On non-disclosure of trade secrets". In this regard, the author apologizes to students of the current work.

[7] https://www.mathworks.com/help/optim/ug/intlinprog.html

[8] Data for calculations were taken from the Enterprise and from the archive of exchange publications.

[9] https://www.mathworks.com/help/deeplearning/ref/fitnet.html

[10] The use of NN is since classical methods of regression analysis did not give positive results.

5. Discussion

Let's look at how the volume of raw materials in the warehouse changed (*Figs. 1, 2*). The behavior of raw materials in the warehouse is stable on average, in contrast to the results of searching for an optimal solution [13], where the maximum value of the total stock of raw materials is reached closer to the 70th day. Moreover, rare emissions show that the state of the levels of raw materials in the warehouse are stable, which allows the enterprise to optimize the management of the warehouse and allocate from it the required warehouse volume as precisely as possible and use the rest for other needs without the need for expansion or vice versa. *Figure 3* shows the production volumes of goods. It can be seen that on average production volumes tend to a maximum of four units. Rare emissions indicate that production is operating stably. When searching for an optimal solution, work [13] shows that production volumes are always equal to 4. The results of the current solution search method are quite close to the optimal one, in terms of production volumes.

The most important indicator remains the value of profit over the entire planning horizon. To estimate the profit trajectory, consider *Fig. 4*. We denote by $opt_m(e)$ the optimal profit value on day m for data sample e.

The behavior of the average values of profit volumes shows that the share of deviation from the optimal solution is not significant and amounts to 0.1964 at the end of the planning horizon. The profit value is growing steadily.

We consider the positive, negative aspects and directions for further development of the current model. The positive aspects of the model include:

1. The conceptual simplicity of the research in terms of modeling – methods for optimizing linear problems have received serious study in science.
2. Considers many important aspects of the forestry industry, for example, the useful volume of raw materials that will arrive at the warehouse, the time of the lot in transit.
3. Allows you to make decisions on each planning day, which corresponds to how the process of production planning and the formation of raw material supply chains occurs.
4. As an analysis of the resulting solution using the example of one of the Primorsky Territory enterprises and exchange data showed, the profit received does not differ much from the optimal value.
5. This model may include other processes of the forestry industry such as logistics, cutting problems and others.
6. As follows from [12, 13, 34], it is necessary to consider changes in railway capacity when forming supply chains. This work takes this feature into account (restrictions (9–10)).
7. The behavior of the values of raw materials in the warehouse is more predictable and stable than with the optimal one, but at the same time there is a payment in profit values.

The negative aspects of the study include the following points:

1. The nature of the β value when calculating the useful volume of raw materials that will reach the warehouse is not known with certainty. It may well be that β is not a parameter, but some function dependent on time or other factors.
2. The work does not consider an important factor in the process of calculating the cost of goods on each individual day or for a period over the entire planning horizon.
3. An important factor that can greatly affect profit – the logistics of finished goods to points of consumption – is not considered in the work, but it is indicated that the model can be modified to a state where it will include this subtask.
4. The key factor in the economy has always been demand. This work does not reflect the influence of demand on the production of goods.
5. Forestry companies often purchase raw materials in a B2B format. It would be useful to also use this tool to diversify the risk of raw material supply chains.
6. Often decisions may have different risk measures. This allows you to get more profit with a competent approach. It would be useful to modify the model

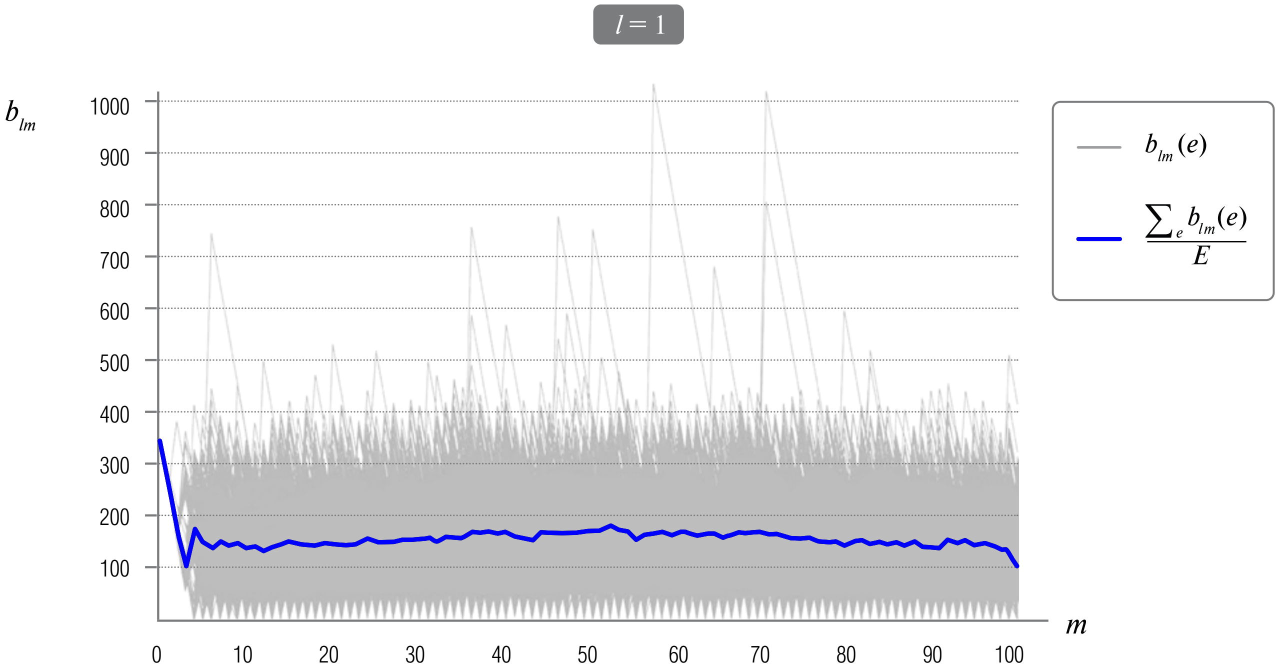


Fig. 1. Visualization of the behavior of the trajectory of raw material inventories of type *l* = 1 in a warehouse over the entire planning horizon.

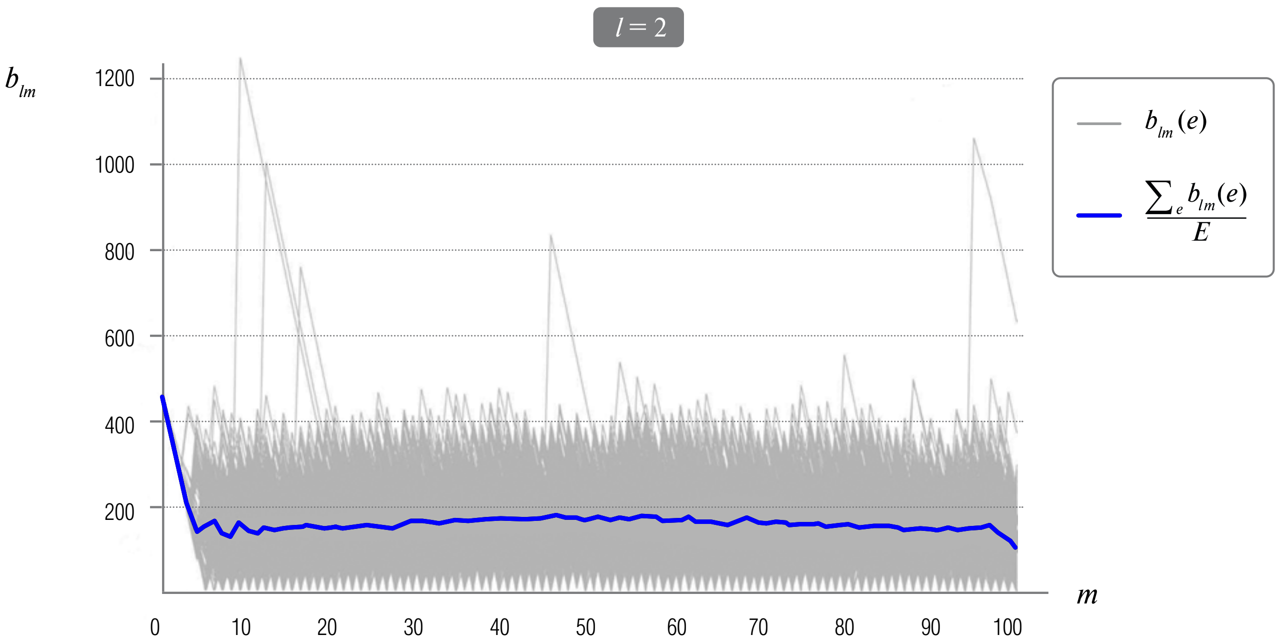


Fig. 2. Visualization of the behavior of the trajectory of raw material inventories of type *l* = 2 in a warehouse over the entire planning horizon.

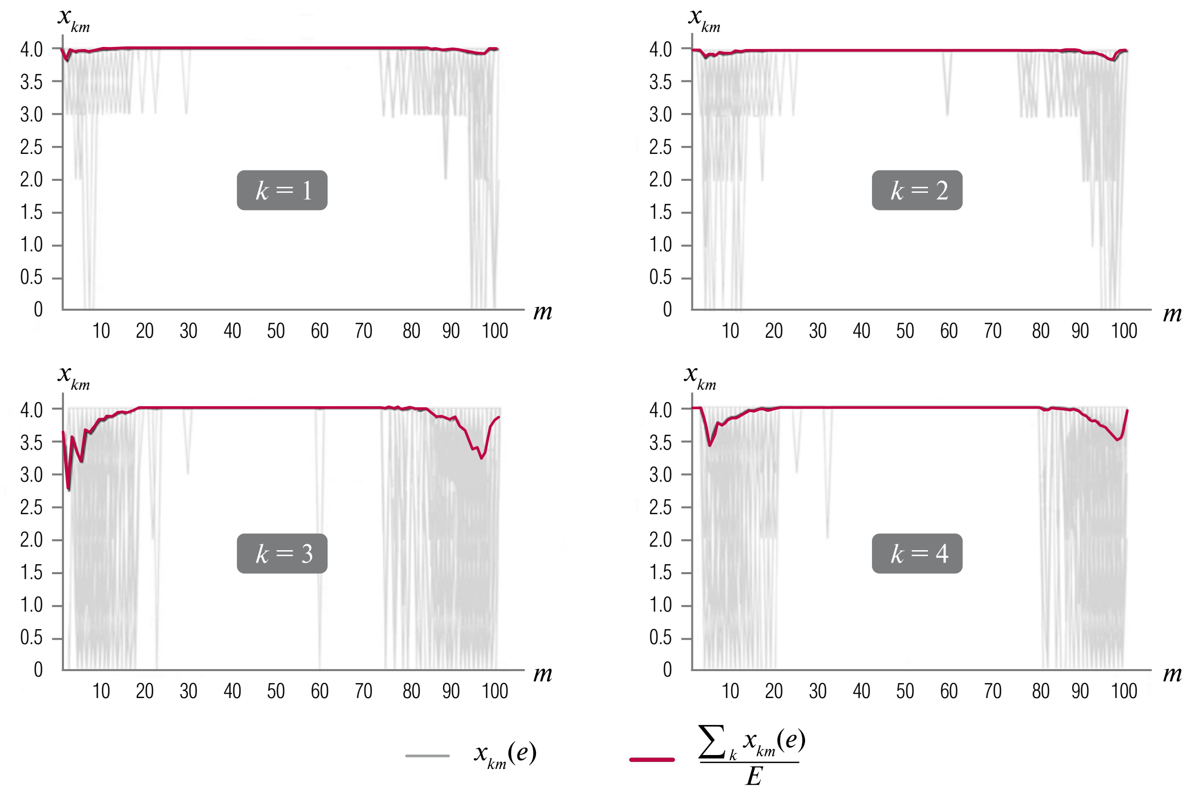


Fig. 3. Visualization of production volumes of goods for each type.

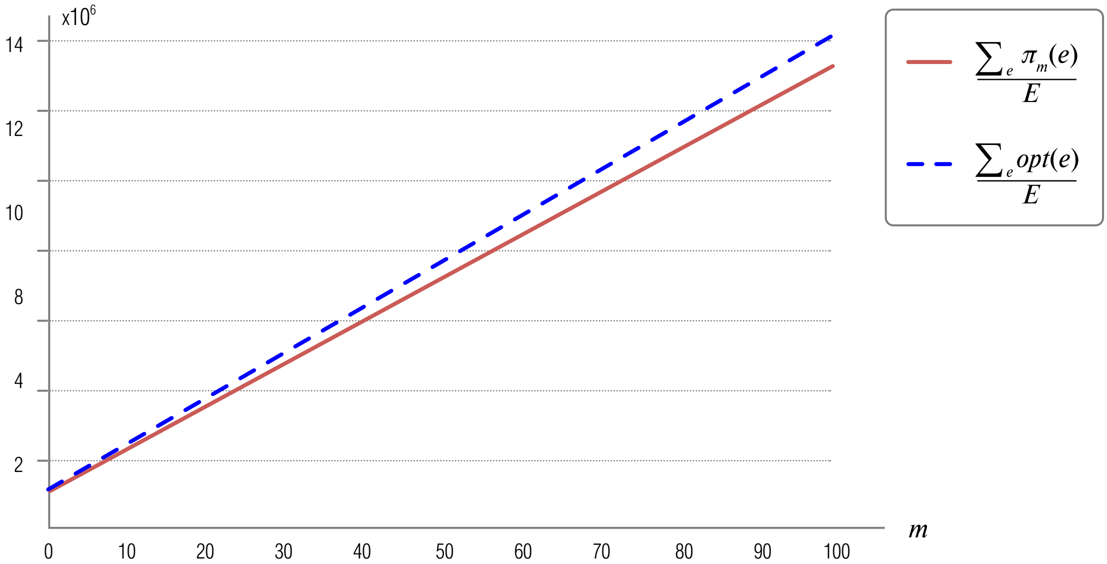


Fig. 4. Visualization of the behavior of the trajectory
of average profit volumes over the entire planning horizon.

so that the decision maker has the opportunity, at a given level of risk, to form supply chains for raw materials.

7. The assessment of parameters ($\tilde{a}_m$, $\tilde{\delta}_m$) responsible for calculating the distance traveled by lots should be calculated not based on the experience of the enterprise (there is a high probability of making an error in planning), but using mathematical methods, for example, neural networks.
8. 2021 and 2022 have shown how important it is to be able to form and rebuild supply chains, including raw materials, under the influence of "black swans" (Black Swan Theory) and/or "rhinos" (Rhino Problem). The current work does not present any considerations on this matter.
9 It is not clear how the model will work if the quality of the estimate of the total volume of raw material stock in the warehouse ($\tilde{b}_m\left(\{V_{imrl}\}_{i\ r}, \{c_{imrl}\}_{i\ r}, m\right)$ and $\tau(m)$) will be low, and will the model show the same results in this situation.

Conclusion

This paper examines a model dedicated to the problem of optimal production management of a timber processing enterprise in matters of calculating production volumes and forming supply chains for raw materials under conditions of uncertainty. The model allows you to maximize the value of profit before tax and is a multi-period linear programming problem characterized by the ability to make decisions simultaneously on both issues under consideration: calculating production volumes and forming supply chains. The results of the model's implementation include the production structure and the sequence of purchasing lots from the commodity exchange, as well as the date of the last receipts at the warehouse. Multi-periodicity is achieved by dividing a task into many smaller ones to solve it on each individual day, just as it happens in enterprises.

The solution search process is a sequential process of solving linear programming problems for making decisions on production volumes and forming supply chains on a daily basis. Every day, the enterprise makes a decision based on estimates of when the goods will arrive at the warehouse and in what volume, where the latter decreases during the delivery process under the pressure of external mechanical factors. To achieve closeness of the solution to the optimal one, it was decided to calculate the regression of the dependence of the target total volume of raw materials in the warehouse on each individual day on the values of the current available lots and the day number. Also for planning, another regression was calculated, which allows us to estimate the dependence of the coefficient of change in the total volume of raw materials in the warehouse on the day number. All regressions were calculated on the data of optimal trajectories of raw material reserves, which were obtained by using one of the well-known models for searching for the optimal solution to the current problem in a complex manner over the entire planning horizon, which took all the necessary played values for the previous period.

The methodology developed for finding a solution was tested using the example of a timber industry enterprise in the Primorsky Territory. Based on the calculations carried out and the solution found, the company's recommendations for cooperation with the Russian Commodity and Raw Materials Exchange were formulated. Analysis of the decision showed that, despite the territorial proximity of the Irkutsk Region to the Primorsky Territory, it is worth paying attention to the purchase of raw materials from the Moscow Region and the Republic of Buryatia. This is explained by many reasons, among which the most important can be identified: sufficient potential in terms of extracted raw materials and a more acceptable pricing policy of companies. A brief analysis of the possible production volumes of each type of product was carried out. Analysis of the solution found demonstrates that the production of most types of goods should be maximum over the entire planning horizon, with rare exceptions.

The positive and negative aspects of the model are given, and ideas for its further development are considered.

In general, it can be argued that the model we developed is effective in finding a solution to the problem. ■

References

1. Kanchana S., Sneha P. (2018) A study on supply chain management in construction projects. *International Research Journal of Engineering and Technology*, vol. 5, pp. 993–996.
2. Banihashemi S.A., Khalilzadeh M., Antucheviciene J., Edalatpanah S.A. (2023) Identifying and prioritizing the challenges and obstacles of the green supply chain management in the construction industry using the fuzzy BWM method. *Buildings*, vol. 13, pp. 38. https://doi.org/10.3390/buildings13010038
3. Cigolini R., Gosling J., Iyer A., Senicheva O. (2022) Supply chain management in construction and engineer-to-order industries. *Production Planning & Control*, vol. 33, pp. 803–810. https://doi.org/10.1080/09537287.2020.1837981
4. Longhui L., Chuan Y., Lirong Q. (2023) Construction supply chain management: A systematic literature review and future development. *Journal of Cleaner Production*, vol. 382, article ID: 135230. https://doi.org/10.1016/j.jclepro.2022.135230
5. Cervený L. Sloup R., Cervená T., Riedl M., Palátová P. (2022) Industry 4.0 as an opportunity and challenge for the furniture industry – A case study. *Sustainability*, vol. 14, article ID: 13325. https://doi.org/10.3390/su142013325
6. Sachan S., Kumar V., Vardhan S., Mittal A., Verma P., Bag S. (2022) Key supply chain strategies for post-COVID-19 recovery: evidence from an Indian smart furniture industry. *International Journal of Emerging Markets*, vol. 18, no. 6, pp. 1378–1396. https://doi.org/10.1108/IJOEM-12-2021-1926
7. Ahmadreza A., Ajang T., Shademan P., Mohammad N. I., Lashgari A. (2022) Sustainable supply chain management and performance in Iran's wooden furniture industry. *Wood Material Science & Engineering*. https://doi.org/10.1080/17480272.2022.2116995
8. Jaehn F., Juopperi R. (2019) A description of supply chain planning problems in the paper industry with literature review. *Asia–Pacific Journal of Operational Research*, vol. 36, no. 01, 1950004. https://doi.org/10.1142/S0217595919500040
9. Mobtaker A., Montecinos J., Ouhimmou M., Rönnqvist M. (2020) Integrated forest harvest planning and road-building model with consideration of economies of scale. *Canadian Journal of Forest Research*, vol. 50(10), pp. 989–1001. https://doi.org/10.1139/cjfr-2019-0380
10. Monti C. A., Gomide L. R., Oliveira R. M., França L. C. (2020) Optimization of wood supply: the forestry routing optimization model. *Anais da Academia Brasileira de Ciências*, vol. 92(3). https://doi.org/10.1590/0001-3765202020200263
11. Rogulin R.S. (2021) Mathematical model for the formation of pricing policy and plan for the production and transport system of a timber industry enterprise. *Business Informatics*, vol. 15, no. 3, pp. 60–77. https://doi.org/10.17323/2587-814X.2021.3.60.77
12. Rogulin R.S. (2020) A model for optimizing the plan for the procurement of raw materials from the regions of Russia by a timber processing complex. *Business Informatics*, vol. 14, no. 4, pp. 19–35. https://doi.org/10.17323/2587-814X.2020.4.19.35
13. Rogulin R.S. Mathematical model for finding the optimal solution to the problem of forming supply chains of raw materials in conditions of uncertainty from the commodity exchange to the warehouse of timber industry enterprises, taking into account production and logistics features. *Journal of Information Technologies and Computing Systems*, 2023 (in press) (in Russian).
14. De S.K., Sana S.S. (2014) A multi-periods production-inventory model with capacity constraints for multi-manufacturers – a global optimality in intuitionistic fuzzy environment. *Applied Mathematical Computations*, vol. 242, pp. 825–841.

15. De S.K., Goswami A., Sana S.S. (2014) An interpolating by pass to Pareto optimality in intuitionistic fuzzy technique for a EOQ model with time sensitive backlogging. *Applied Mathematical Computations*, vol. 230, pp. 664–674.

16. Chakraborty D., Jana D.K., Roy T.K. (2014) A new approach to solve intuitionistic fuzzy optimization problem using possibility, necessity, and credibility measures. *International Journal of Engineering Mathematics*, pp. 1–12. https://doi.org/10.1155/2014/593185

17. Ahmadini A.A.H., Modibbo U.M., Shaikh A.A., Ali I. (2021) Multi-objective optimization modelling of sustainable green supply chain in inventory and production management. *Alexandria Engineering Journal*, vol. 60, pp. 5129–5146. https://doi.org/10.1016/j.aej.2021.03.075.

18. Chakrabortty S., Pal M., Nayak P.K. (2012) Intuitionistic fuzzy optimization technique for Pareto optimal solution of manufacturing inventory models with shortages. *European Journal of Operations Research*, vol. 228(2), pp. 381–387.

19. Garg H., Rani M. (2013) An approach for reliability analysis of industrial systems using PSO and IFS technique. *ISA Transactions*, vol. 52(6), pp. 701–710.

20. Garg H., Rani M., Sharma S., Vishwakarma Y. (2014) Bi-objective optimization of the reliability-redundancy allocation problem for series-parallel system. *Journal of Manufacturing Systems*, vol. 33(3), pp. 335–347. https://doi.org/10.1016/j.jmsy.2014.02.008

21. Garg H., Rani M., Sharma S., Vishwakarma Y. (2014) Intuitionistic fuzzy optimization technique for solving multi-objective reliability optimization problems in interval environment. *Expert Systems with Applications*, vol. 41(7), pp. 3157–3167. https://doi.org/10.1016/j.eswa.2013.11.014

22. Roy B.V., Bertsekas D.P., North L. A neuro-dynamic programming approach to admission control in ATM networks. Proceedings of the *IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, 21–24 April 1997.*

23. Kleywegt A.J., Non V.S., Savelsbergh M.W. (2004) Dynamic programming approximations for a stochastic inventory routing problem. *Transportation Science*, vol. 38, pp. 42–70.

24. Topaloglu H., Kunnumkal S. (2006) Approximate dynamic programming methods for an inventory allocation problem under uncertainty. *Naval Research Logistics*, vol. 53, pp. 822–841. https://doi.org/10.1002/nav.20164

25. Kunnumkal S., Topaloglu H. (2008) Using stochastic approximation methods to compute optimal base-stock levels in inventory control problems. *Operations Research*, vol. 56, pp. 646–664.

26. Cimen M., Kirkbride C. Approximate dynamic programming algorithms for multidimensional inventory optimization problems. Proceedings of the *7th IFAC Conference on Manufacturing, Modeling, Management, and Control, Saint Petersburg, Russia, 19–21 June 2013*, vol. 46, pp. 2015–2020.

27. Perez H.D., Hubbs C.D., Li C., Grossmann I.E. (2021) Algorithmic Approaches to Inventory Management Optimization. *Processes*, vol. 9, article ID: 102. https://doi.org/10.3390/pr9010102

28. Mortazavi A., Khamseh A.A., Azimi P. (2015) Designing of an intelligent self-adaptive model for supply chain ordering management system. *Engineering Applications of Artificial Intelligence*, vol. 37, pp. 207–220. https://doi.org/10.1016/j.engappai.2014.09.004

29. Luo L., O'Hehir J., Regan C.M., Meng L., Connor J.D., Chow C.W.K. (2021) An integrated strategic and tactical optimization model for forest supply chain planning. *Forest Policy and Economics*, vol. 131, article ID: 102571. https://doi.org/10.1016/j.forpol.2021.102571

30. Kara A., Dogan I. (2018) Reinforcement learning approaches for specifying ordering policies of perishable inventory systems. *Expert Systems with Applications*, vol. 91, pp. 150–158.

31. Prajapati D., Chan F.T.S., Chelladurai H., Lakshay L., Pratap S. (2022) An Internet of Things embedded sustainable supply chain management of B2B E-Commerce. *Sustainability*, vol. 14, article ID: 5066. https://doi.org/10.3390/su14095066

32. Wang S., Fang Z., Wu D. (2022) Internet of things-enabled tourism economic data analysis and supply chain modeling. *Technological and Economic Development of Economy*, pp. 1–18. https://doi.org/10.3846/tede.2022.17120

33. Li J., Zhang R., Jin Y., Zhang H. (2022) Optimal path of internet of things service in supply chain management based on machine learning algorithms. *Computational Intelligence and Neuroscience*, article ID 4844993. https://doi.org/10.1155/2022/4844993

34. Vinogradova M., Rogulin R., Ermakova M., Okhrimenko I. (2021) Assessing the sources of uncertainty in supply chain management. *Strategic Change*, vol. 30(5), pp. 453–460. https://doi.org/10.1002/jsc.2465

About the author

Rodion S. Rogulin

Cand. Sci. (Econ.);

Associated Professor, Department of Mathematics and Modeling, Vladivostok State University, 41, Gogolya Street, Vladivostok 690014, Russia;

Senior Lecturer, Software and Artificial Intelligence Department, Far Eastern Federal University, 10, Ajax Village, Russky Island, Primorskiy Territory 690922, Russia;

E-mail: rafassiaofusa@mail.ru

ORCID: 0000-0002-3235-6429

DOI: 10.17323/2587-814X.2023.4.57.72

The influence of the breadth of the tech stack on the result of the digital product: A view through the theory of absorption capacity

Ivan V. Solovyov
E-mail: solovjev.iw@yandex.ru, isolovev@hse.ru

Viatcheslav A. Semenikhin
E-mail: vsemenikhin@planear.ru, vasemenikhin@hse.ru

Sergey P. Kushch
E-mail: skushch@hse.ru

Graduate School of Business, HSE University
Address: 26–28, Shabolovka Str., Moscow 119049, Russia

Abstract

In today's economy based on knowledge and innovation, the development of absorptive capacity by companies is a critical aspect of business competitiveness. In this study, the tech stack of sites is considered as a specific, measurable part of the digital and innovative development of a company. In literature to date, there is no clear answer to which technologies and in what quantity should be included in the tech stack.

From the point of view of assessing the tech stack, mainly qualitative methods are proposed that are quite resource intensive. Accordingly, the purpose of this study is to determine the impact of the technologies used by the characteristics of quantity, uniqueness and popularity in the tech stack of the product on the result (the absence of critical errors); as well as in developing a quantitative approach for assessing the impact of the technologies used on the result of a digital product. The quantitative approach was developed and conceptualized based on previous literature, tested on 12 sites of large Russian banks, including 12 main domains and 595 subdomains. An analysis of a study of 216 online applications for banking products showed a positive relationship between the share of unique technologies in the bank's visible tech stack and the number of errors, as well as a negative relationship between the share of popular technologies in the stack and errors. This study expands the discussion on the development of absorptive capacity, contributes to the understanding of the limitations of absorptive capacity of companies and proposes a quantitative approach for auditing the operational tech stack of companies' websites.

Keywords: tech stack, innovation, absorption capacity, digital marketing, website

Citation: Solovyov I.V., Semenikhin V.A., Kushch S.P. (2023) The influence of the breadth of the tech stack on the result of the digital product: A view through the theory of absorption capacity. *Business Informatics*, vol. 17, no. 4, pp. 57–72. DOI: 10.17323/2587-814X.2023.4.57.72

Introduction

Business in the modern world is highly dependent on knowledge and innovative development. Many researchers note a company's ability to innovate and make current changes as one of the main factors of competitiveness and prospects for its development (for example: [1, 2]). However, knowledge leading to current business transformations is mainly generated outside a specific company, in uncontrolled dynamic systems that include a large number of actors [3], which raises issues related to the identification, assimilation and use of external knowledge.

In 1990, Cohen and Levinthal [4] proposed the concept of absorptive capacity of organizations, which includes the ability to recognize the value of new information, assimilate it and apply it for business purposes. Over the past decades, this concept has received serious development: a number of researchers have focused on the development of conceptualization and operationalization of this concept [5, 6]; other researchers have turned to the study of internal company factors and external environmental factors affecting the absorptive capacity of companies [3, 7–9]. Another significant direction in the development of this concept was the study of the effects on the company from the development of absorptive capacity.

The knowledge obtained in the framework of the latest direction of research is currently quite fragmentary, sometimes contradictory. So, despite the fact that a number of studies confirm a positive relationship between the introduction of innovations in a company and the results demonstrated [10–14], there are also studies that indicate that the relationship may also be negative: the desire for innovation can lead to more risky, complex and less linear processes [15] and (potentially) more asymmetrical returns [16]. Other scholars argue that firms with high innovative activity suffer from low collateral assets and long and uncertain payback periods [17, 18].

In 2022, Lehmann, Menter and Wirsching found that the relationship between firm productivity and absorptive capacity has an inverted U-shape [9], indicating that companies have a certain optimum point, but after that is crossed an increase in knowledge absorption, an increase in investment in R&D is not practical. However, no convincing proposals have yet been presented as to how to determine this very optimum point; how to identify useful and useless innovations; how to evaluate your own company from the point of view of sufficient innovation saturation and avoid negative influence.

This study examines the tech stack of banking websites as a specific, measurable part of a company's digital and innovative development. Today it is impossible to imagine an organization without an online presence in the form of a website, landing pages for promoting specific products or services, and online forms for filling out applications. The more a company interacts with its customers in digital channels, the more attention it pays to online technologies and the tech stack.

With the growing diversification of available technologies, the expansion of functionality, the increase in the number of suppliers of such tools, as well as changes in legislation, for example, on privacy protection [19], the tech stack has become an extremely promising object of analysis in terms of its optimization and the use of certain technologies.

By including a large number of technological tools in their own stack (increasing its breadth), banks, on the one hand, demonstrate a high level of development of absorptive capacity; on the other hand, they increase the risks of their joint "unassimilation," which can lead to technical errors and failures in the operation of the digital product. Researchers believe that auditing the current tech stack not only allows you to get rid of unnecessary technologies, but also contributes to the development of a more informed approach to choosing technologies in the future [20, 21].

However, to study and audit the tech stack, researchers and practitioners mainly propose a qualitative approach, which involves analyzing each technology separately for the need for its use [20]. Despite the fact that this approach has certain advantages, it nevertheless requires serious human and time resources.

In accordance with the identified gaps, the purpose of this study is to determine the impact of the technologies used by the characteristics of quantity, uniqueness and popularity in the product tech stack on the result; in developing a quantitative approach to assessing the impact of the technologies used on the quality of the product.

The following research questions were formulated:

1. How does increasing the breadth of the tech stack (inclusion of a large number of technologies) affect the quality of the product (number of errors)?
2. How does the use of unique technologies affect the quality of the product (the number of errors)?
3. How does the use of popular technologies affect the quality of the product (the number of errors)?

This study presents and tests a quantitative approach to assessing the performance of a product such as a company website (main domain and all subdomains) based on an analysis of the tech stack. A positive relationship was found between the share of unique technologies and the number of errors, as well as a negative relationship between the share of popular technologies and the number of errors.

This study contributes to the development of absorptive capacity theory, in particular in the direction of studying the results of the company obtained from the development of absorptive capacity expressed in the use of a wide range of external products, and also offers a concrete practical tool for assessing the effectiveness of adding a large number of products to its own tech stack from unique technologies and popular technologies.

1. Absorptive capacity

In the seminal work of Cohen and Levinthal (1990), absorptive capacity is defined as "the ability to recognize the value of new information, assimilate it and apply it to commercial purposes" [4]. One of the most important assumptions of this concept is that the ability to find and use external knowledge contributes to the development of the innovative potential of companies, which is especially important in a knowledge-based economy [3]. Cohen and Levinthal (1990) called absorptive capacity generators a company's investments in R&D, the company's production activities and investments directly in absorptive capacity (for example, through personnel training) [4].

Further development of this concept was taken up by many researchers, who also tried to rethink it and create a more precise conceptual framework. In particular, Zahra and George [5] identified four capabilities that collectively represent a company's absorptive capacity: acquisition, assimilation, transformation and exploitation. Moreover, in response to the question of whether all acquired knowledge can be assimilated and used, researchers have proposed dividing absorptive capacity into two subcategories: potential absorptive capacity, which includes the processes of acquiring and assimilating knowledge; realized absorptive capacity, which includes the processes of transformation and exploitation of knowledge [5].

The division of absorptive capacity into potential and realized kinds has become a natural consequence of the problem of knowledge acquired but not used by companies for various reasons [8]. Among the factors that have a direct impact on a company's absorptive capacity are internal factors: previous knowledge base [4, 6, 7, 9], absorptive capacity, employee competencies, company size, investments in R&D [4, 6], organizational structures [6, 7] and others; external: knowledge environment, company position in knowledge networks [6].

Also, one of the most important questions about the use of external knowledge by companies is whether the innovative potential of companies will grow indefinitely with the constant acquisition of new knowledge. In 2022, Lehmann, Menter and Wirsching found that the relationship between firm productivity and absorptive capacity has an inverted U-shape [9], indicating that companies have a certain optimum point, and after they cross that an increase in knowledge absorption, an increase in investment in R&D is not practical. However, methods for estimating the optimum, as well as the reasons for the decrease in efficiency after crossing it, remain insufficiently studied.

Thus, the assumption that not all acquired knowledge can be absorbed and implemented by companies, as well as the assumption that "more is not better," are key prerequisites for conducting this study and developing a quantitative approach to assessing product performance depending on the number of built-in innovation.

2. Tech stack

A tech stack is a set of technologies on the basis of which digital applications and websites are developed. Various digital tools, databases, programming languages, etc. can be integrated into the tech stack [22]. With the growing diversification of available technologies, the expansion of functionality, the increase in the number of suppliers of such tools, as well as changes in legislation, for example, on privacy protection [19], the tech stack has become an extremely promising object of analysis in terms of its optimization and the use of certain technologies.

The motivation for increasing the complexity of one's own tech stack and integrating a large number of digital solutions into it, in addition to functional benefits, is also the growing demand of customers who expect to see increasingly automated and high-tech solutions to their problems. At the same time, a well-designed tech stack can be a source of competitive advantage [21].

All this encourages companies to constantly study the emergence of new technologies and determine priorities for investing in the company's tech stack [23]. In addition to searching for new technologies, modern researchers emphasize the need to conduct regular audits of their tech stacks for redundancy. Moreover, researchers also believe that an audit of the current tech stack allows you not only to get rid of unnecessary technologies, but also contributes to the development of a more informed approach to the choice of technologies in the future [20, 21]. This also corresponds to the provisions of the development of absorptive capacity and the need to have prior knowledge for the correct assimilation of new ones. In addition, the capabilities of a tech stack significantly depend on the team that is involved in its development and optimization [21].

One of the most important problems that can arise from the incorrect use of technology or the use of untested technologies for digital products is the emergence of "critical" errors that prevent the achievement of key product goals.

For example, the main indicator of the effectiveness of a landing page is the conversion from a visitor to those who completed an application [24], and the interaction of consumers with the company's website as a whole is an important part of the company's relationship with its audience [25]. A landing page is a complex product that can have an arbitrarily complex front-end structure (everything that the browser can read, display and/or run) and a back-end, and behave differently depending on the environment in which it is tested, such as different phone operating system models, different brands, and different browser versions. All this, on the one hand, actualizes the desire of companies to introduce more and more new innovations into their own tech stack, but, on the other hand, it leads to an increase in the risks of "critical errors."

In this study, the result of the landing page tech stack is assessed precisely by identifying the presence or absence of "critical" errors, that is, those that do not allow users to complete the filling out of the questionnaire (for example, SMS does not arrive to confirm data, or the "next" button is not pressed to fill out a form, or after going to the State Services portal and obtaining consent to use the data, this data is not saved and must be re-entered manually, etc.). Based on the above, we formulate the first research question:

– How does increasing the number of technologies in the tech stack affect the results of products (the number of errors)?

The use of newly emerging or simply rare technologies for the market requires time, human resources and specialization, and, ultimately, R&D costs to absorb and assimilate them. According to this study, we seek to determine whether there is a relationship between the number of unique technologies (which no one except this company or firm uses) and the number of errors; is there a negative relationship between the share of popular technologies (which are used by more than half of the agents under study) and the number of errors. Accordingly, two other research questions:

– How does the use of unique technologies affect the results of products (the number of errors)?

– How does the use of popular technologies affect the results of products (the number of errors)?

To study the tech stack, researchers and practitioners mainly offer a qualitative approach which is associated with the analysis of each technology separately for the need for its use [20]. Despite the fact that this approach has certain advantages over quantitative research, it nevertheless requires serious human and time resources.

This study proposes an original quantitative approach to conducting an audit of a website's tech stack, allowing one to assess the state of one's own tech stack in comparison with similar companies, as well as identify technologies that could potentially influence the increase in the number of errors.

3. Quantitative approach

In order to assess the number of technologies used in the tech stack of banking websites, we chose the well-known BuiltWith service, which collects and classifies the majority of known technologies on all registered domains since 2002, based on signals from the websites themselves about the use of a particular technology.

The choice of industry for the study was based on review in the most developed companies in terms of technology implementation. Banks were chosen as research objects, since the Russian financial sector is one of the leaders in digital transformation. "According to calculations by ISIEZ HSE University according to Rosstat, the digitalization index of the domestic financial sector at the end of 2019 reached a value of 34 and was second only to industry (with an indicator of 36)" [26, p. 159], while "the internal costs of financial sector organizations for the creation, distribution and use of digital technologies and related products and services amounted to 380.2 billion rubles, which corresponds to 8.9% of the gross added value of the sector, leaving all other sectors of the economy and social sphere far behind in these most important indicators" [26, pp. 159–160]. According to the results of the FINIX study by Yakov and Partners (the former Russian division of McKinsey) in March 2023, "large Russian banks successfully survived the shocks of 2022 and maintained global leadership in terms of digitalization" [27]. Banks are therefore a particularly interesting sector from a tech stack research perspective.

Banks were selected for the study based on asset indicators [28], after which requests were sent to check the technologies used with the help of the above-mentioned Built With service. Some banks did not provide permission to the service robots to record technology; therefore, only those banks for which data was available were included in the final list.

The collected lists of technologies found on the main sites and subdomains were transferred to Excel spreadsheets and analyzed. The number of technologies was calculated for each of the standard rubricators of the service, then the total number of technologies used by the bank was calculated, as well as the number of non-repeating technologies in the stack of each bank, the share of unique technologies in the stack (the number of technologies, that is, used by only one bank from the list to the number of non-repeating technologies used by this bank), "rare" technologies in the stack (the number of technologies used by only 1–2 banks from the list to the number of non-repeating technologies used by this bank), "popular" technologies in the stack (the number of technologies used by the majority, that is, more than 6 banks from list to the number of non-repetitive technologies used by this bank) [29].

To assess the performance of the sites, the process of filling out online applications for banking products to obtain bank approval was studied. The screen of a mobile phone or computer was recorded. Next, for the purposes of our study, we counted the cases of critical errors occurring when filling out the questionnaire (which should not have occurred). We called critical errors when the user could not continue filling out the form (for example, the SMS does not arrive to confirm the data, or the "next" button is not pressed to fill out the form, or after going to the State Services portal and obtaining consent to use the data, this data is not saved and they must be entered again manually, etc.). In 2021–2022, materials were publicly published based on this study, which took into account the number of errors recorded by researchers. Most of the banks studied had signed agreements providing for a detailed review of the application process and review of errors. There has not been a single precedent when a bank challenged a factual error recorded and recorded on the screen of a smartphone or computer. Links to public materials are provided in Appendix.

A non-parametric Spearman test was then applied to look for a correlation between the above tech stack metrics and the percentage of errors encountered when filling out online questionnaires.

4. Results

At the time of the study on June 1, 2022, 12 domains and 595 subdomains of the studied banks were available for verification (*Table 1*).

After grouping by technology type in accordance with the classifier proposed by the BuiltWith service, 529 non-repeating technologies were discovered, collected in 25 groups, which were used by banks 2189 times (*Table 2*). The leaders in terms of the number of uses by banks include the following technologies: JavaScript Libraries and Functions (622 times); Analytics and Tracking (209 times); Frameworks (190 times). The technologies are used the greatest number of times in Tinkoff, Alfa-Bank, Otkritie Bank, the least in Rosbank, SMP Bank and Gazprombank.

Table 1.

Researched banks, domains and subdomains

Bank name	Domain name	Number of subdomains examined
Alfa-Bank	ALFABANK.RU	100
BSPB	BSPB.RU	25
Gazprombank	GAZPROMBANK.RU	25
MKB	MKB.RU	64
MTS Bank	MTSBANK.RU	102
Otkritie Bank	OPEN.RU	52
Raiffeisenbank	RAIFFEISEN.RU	57
Rosbank	ROSBANK.RU	36
Rosselkhozbank	RSHB.RU	38
SMP Bank	SMPBANK.RU	29
Sovcombank	SOVCOMBANK.RU	30
Tinkoff Bank	TINKOFF.RU	37
Total	**12**	**595**

Table 3 for each bank calculates the total number of unique (used in only one bank), rare (used in two banks or less) and popular (used by more than six banks) technologies.

The largest number of unique technologies was discovered in Alfa-Bank (54 technologies), which is 22.6%, and Tinkoff, which is 30.4%.

The number of detected and recorded errors when filling out online applications for banking products was 25 out of 216 surveyed questionnaires (12%). The frequency of errors per questionnaire for each bank is shown in *Table 4*.

To detect the relationship between tech stack indicators and errors, the following indicators were calculated by bank:

a) the total number of technologies used;

b) the number of non-repetitive technologies;

c) the share of unique technologies in the bank's stack (unique – that is, used only by this bank from the list of studied banks);

d) the share of "rare" technologies (that is, used in 1–2 of the banks studied);

e) the share of "popular" technologies (that is, used in more than half of the banks studied).

Table 2.

Identified technologies divided into groups within the Builtwith service

Technologies \ Banks and their main web domains	Alfa-Bank ALFABANK.RU	BSPB B SPB.RU	Gazprombank GAZPROMBANK.RU	MKBank MKB.RU	MTS Bank MTSBANK.RU	Otkritie Bank OPEN.RU	Raiffeisenbank RAIFFEISEN.RU	Rosbank ROSBANK.RU	Rosselkhozbank RSHB.RU	SMP Bank SMPBANK.RU	Sovcombank SOVCOMBANK.RU	Tinkoff Bank TINKOFF.RU	Total
Advertising	23	6	5	6	7	8	3	5	3	4	6	9	**85**
Analytics and Tracking	28	15	10	10	18	27	16	17	11	7	17	33	**209**
Audio / Video Media	2		1	1	2	5	4	1			4		**20**
Content Delivery Network	14	6	1	4	5	10	8	2	8		3	3	**64**
Content Management System	10	2	2		9	3	2		2	3		4	**37**
Copyright	2	3	1	3	1	3	3	1		1			**18**
e-Commerce	1			1					1			1	**4**
Email Hosting Providers	6	2	4	2	5	4	4	2	2	2	4	9	**46**
Frameworks	36	10	3	45	18	14	6	6	6	5	18	23	**190**
JavaScript Libraries and Functions	81	49	37	81	48	66	55	22	39	37	45	62	**622**
Language	1	1	1	1	1	4	1	1	1	1	3	1	**17**
Mapping	1	1	2	1	1		2		1			1	**10**
Mobile	30	9	8	14	11	13	27	6	8	10	7	9	**152**
Name Server			2			1				1		4	**8**
Operating Systems and Servers	6	4	2		2	2	2	1	1	1	3	2	**26**
Payment	7	4	4	8	6	7	3	6	4	2	6	13	**70**
Robots.txt	8	4	1	1	2	7	1	3	2	2	6	14	**51**
SSL Certificates	25	8	8	18	18	30	20	4	6	3	6	32	**178**
Syndication Techniques	3					1						2	**6**
Verified CDN	1					1							**2**
Verified Link									3		1	1	**5**

Technologies \ Banks and their main web domains	Alfa-Bank ALFABANK.RU	BSPB B SPB.RU	Gazprombank GAZPROMBANK.RU	MKBank MKB.RU	MTS Bank MTSBANK.RU	Otkritie Bank OPEN.RU	Raiffeisenbank RAIFFEISEN.RU	Rosbank ROSBANK.RU	Rosselkhozbank RSHB.RU	SMP Bank SMPBANK.RU	Sovcombank SOVCOMBANK.RU	Tinkoff Bank TINKOFF.RU	Total
Web Hosting Providers	9	4	3	1	6	4	4				1	50	**82**
Web Master Registration	1	2		3	1	2	4	1	2		2	2	**20**
Web Servers	13	16	8	1	15	19	9	1	4	8	6	18	**118**
Widgets	20	21	5	14	11	17	8	10	6	3	14	20	**149**
Number of technologies used	328	167	108	215	187	248	182	89	110	90	152	313	**2189**
Number of non-repeating technologies used	239	134	96	123	140	172	126	85	107	73	136	168	**529**

Table 3.

Analysis of technologies in tech stacks of bank websites

Banks and their main webdomains	Number of unique technologies	Number of rare technologies	Number of popular technologies	Share of unique technologies	Share of rare technologies	Share of popular technologies
Alfa-Bank ALFABANK.RU	54	96	71	22.6%	40.2%	29.7%
BSPB BSPB.RU	22	33	68	16.4%	24.6%	50.7%
Gazprombank GAZPROMBANK.RU	7	18	49	7.3%	18.8%	51.0%
MKBank MKB.RU	20	30	55	16.3%	24.4%	44.7%
MTS Bank MTSBANK.RU	15	32	67	10.7%	22.9%	47.9%
Otkritie Bank OPEN.RU	25	44	68	14.5%	25.6%	39.5%
Raiffeisenbank RAIFFEISEN.RU	12	25	57	9.5%	19.8%	45.2%
Rosbank ROSBANK.RU	4	8	58	4.7%	9.4%	68.2%
Rosselkhozbank RSHB.RU	19	25	54	17.8%	23.4%	50.5%
SMP Bank SMPBANK.RU	6	12	43	8.2%	16.4%	58.9%
Sovcombank SOVCOMBANK.RU	22	34	71	16.2%	25.0%	52.2%
Tinkoff Bank TINKOFF.RU	51	76	61	30.4%	45.2%	36.3%

Table 4.

Frequency of errors detected during the study of online questionnaires of banks

Domain	Number of applications filled	Number of errors	Frequency of errors
ALFABANK.RU	21	1	0.05
BSPB.RU	8	2	0.25
GAZPROMBANK.RU	21	1	0.05
MKB.RU	15	5	0.33
MTSBANK.RU	17	1	0.06
OPEN.RU	21	4	0.19
RAIFFEISEN.RU	21	4	0.19
ROSBANK.RU	21	0	0.00
RSHB.RU	21	5	0.24
SMPBANK.RU	8	1	0.13
SOVCOMBANK.RU	21	0	0.00
TINKOFF.RU	21	1	0.05
Total	**216**	**25**	**12%**

The results obtained are shown graphically in *Fig. 1*.

Based on the data presented in *Fig. 1*, we can conclude that the population of banks is heterogeneous in terms of rare or unique technologies; two banks with a higher number of unique technologies stand out significantly (see *Table 4*): this is Alfa-Bank (54 unique technologies) and the bank Tinkoff (51). They also stand out in terms of "rare technologies" – there are 96 of them at Alfa-Bank and 76 at Tinkoff Bank. The shares of "rare" (40% and 45%, respectively) and unique technologies (23% and 30%) are maximum for the stacks of each of these banks in comparison with the rest of the ones studied. At the same time, the share of technologies used by most banks in the stack of these banks is minimal for the sample (30% and 36%). That is, both banks stand out noticeably from the general population in terms of stack indicators.

At the next stage of analysis, to identify the relationship, the nonparametric Spearman rank correlation test was calculated, which turned out to be insignificant for all selected groups within the entire population of banks ($n = 12$). However, after separating the two specified banks – Tinkoff and Alfa Bank – from the population (in terms of the number or share of unique technologies in the stack), for the rest of the banks ($n = 10$) a significant inverse correlation (p less than

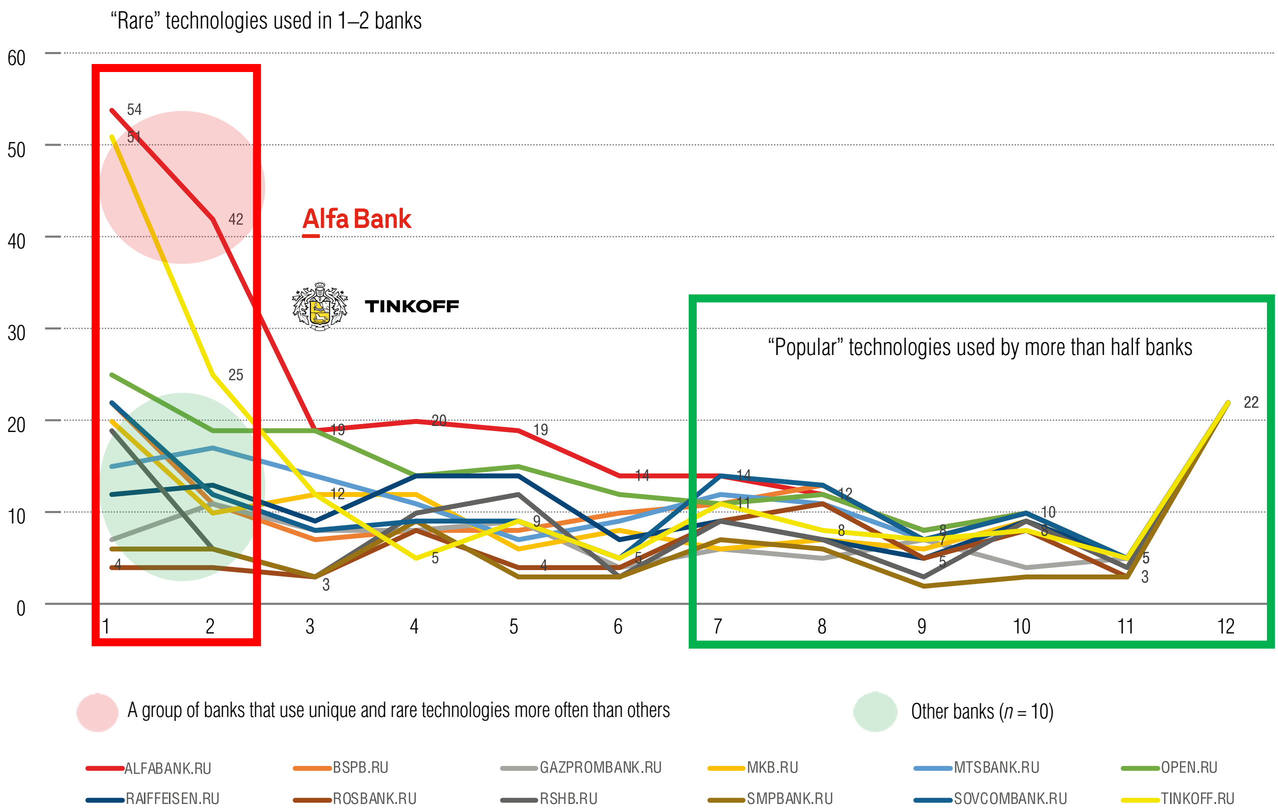


Fig. 1. Distribution of unique technologies by frequency of occurrence in the banks studied ($n = 12$):
x-axis – how many banks use the same technology,
y-axis – the number of such technologies in the bank.

0.05) was found between the proportion of errors in the questionnaires and the share of popular technologies in the bank's stack, and the same significant correlation between the share of errors and the share of unique technologies in the bank's stack (*Table 5*).

According to the results obtained, no statistically significant correlation was found between the increase in the number of technologies in the tech stack and the result of the product (number of errors). A similar result was obtained after removing two outlier banks ($n = 10$), and although the Spearman indicator increased, the correlation did not reach the significance level of 0.05.

The indicator of the share of unique technologies in the tech stack also did not show a significant correlation with the result of product performance (number of errors) for all banks ($n = 12$) but showed a significant correlation ($p < 0.05$) after removing two outstanding banks from the sample ($n = 10$). That is, the greater the share of unique technologies in the stack, the more likely it is that errors will appear on the landing page.

The share of popular technologies in the stack has an inverse correlation with the number of errors ($p < 0.05$) for the sample after excluding Alfa-Bank and Tinkoff Bank ($n = 10$) and may be a good indicator. In contrast to the uniqueness of the stack, we see that for most of the banks studied, the use of popular technologies is inversely related to the indicator of the quality of the site (the number of errors), and at the same time

Table 5.

Indicators of technology distribution by studied domains and subdomains of banks, surveyed questionnaires, detected errors and correlations by group

Domain	Number of technologies used	Number of non-repeating technologies used	Share of rare technologies (2–)	Share of unique technologies (<2)	Share of popular technologies (7+)	Number of applications studied	Errors detected	Errors as a share of questionnaires
ALFABANK.RU	328	239	40%	23%	30%	21	**1**	0.05
BSPB.RU	167	134	25%	16%	51%	8	**2**	0.25
GAZPROMBANK.RU	108	96	19%	7%	51%	21	**1**	0.05
MKB.RU	215	123	24%	16%	45%	15	**5**	0.33
MTSBANK.RU	187	140	23%	11%	48%	17	**1**	0.06
OPEN.RU	248	172	26%	15%	40%	21	**4**	0.19
RAIFFEISEN.RU	182	126	20%	10%	45%	21	**4**	0.19
ROSBANK.RU	89	85	9%	5%	68%	21	**0**	0.00
RSHB.RU	110	107	23%	18%	50%	21	**5**	0.24
SMPBANK.RU	90	73	16%	8%	59%	8	**1**	0.13
SOVCOMBANK.RU	152	136	25%	16%	52%	21	**0**	0.00
TINKOFF.RU	313	168	45%	30%	36%	21	**1**	0.05
Spearman correlation ($n = 12$)	0.14	–0.10	0.01	0.23	–0.22	216	25	12%
Spearman correlation, without Alfa Bank and Tinkoff Bank ($n = 10$)	0.51	0.11	0.37	0.69	–0.63			
Significance ($n = 10$)				$p < 0,05$	$p < 0,05$			

can bring obvious benefits for the development and support of complex digital products without the use of rare technologies and rare specialists, obtaining possible savings in payroll and reducing the diversity of supported competencies in product development and testing.

5. Discussion

This study attempts to assess the impact of technologies included in the tech stack of banking websites on the performance of a given digital product, as an example of the impact of absorptive capacity.

Having studied the tech stack of sites for the number of technologies used in the tech stack and for the number of errors, a connection between them was not found, which is most likely due to the complex nature of technological development and the unique approaches of each individual bank to the selection and integration of technologies, which is consistent with previous studies, indicating that the success of innovation absorption depends on factors such as the previous knowledge base [4, 6, 7, 9], absorptive capacity, employee competencies, company size, R&D investments [7, 14], organizational structures [4, 6], etc.

However, a moderate positive relationship was found between the share of unique technologies and the number of errors for ten banks, after excluding two outliers from the analysis. That is, the greater the share of unique technologies in the stack, the more likely it is that errors will appear on the landing page. In interpreting the result obtained, we draw attention to the fact that "stand-out banks" are an example of banks that have relied on the uniqueness of the stack they use (judging by the indicators of a high share of unique and rare technologies and a low share of "popular" technologies). Also, these banks are leaders in the main technology ratings in the Russian market, such as MarksWebb [30], Banking Review [31], FrankRG [32], these banks have proven themselves in working with their unique stack and demonstrate good business results. For banks that have not yet defined their unique stack, it is probably worth paying attention to the use of a quantitative approach to assessing the tech stack, as a prognostic indicator that allows you to quickly assess the degree of uniqueness of the technologies used and possibly take measures to qualitatively reduce the unjustified diversity of technologies to reduce the number of errors. Additional benefits from reducing the number of unique technologies may include an advantage in the selection and recruitment of personnel (there is no need to search and test rare specialists), launch speed (reducing the stages of familiarization with the technology and maintaining competencies in the use of technologies unique to the market), and the absence of an "overpayment for the uniqueness" of a specialist, reducing the risk of finding a replacement for a unique specialist, the unpredictability of the behavior of new exotic technologies, etc.

A moderate negative relationship was also found between the share of popular technologies and the number of errors for ten banks, after excluding two outliers from the analysis. In addition to the finding that the number of unique technologies leads to an increase in the number of errors, it was found that for the majority of banks studied, the use of popular technologies is inversely correlated with the indicator of the quality of website performance (number of errors), and at the same time can provide obvious benefits in terms of developing and maintaining complex software products without the use of rare technologies and rare specialists, obtaining possible savings in the wage fund and reducing the variety of supported competencies in product development and testing.

The results obtained during the study confirm the current discussion about the need to audit companies' tech stacks [20, 21]. Using the proposed quantitative approach will allow companies to determine the degree of uniqueness and popularity of a tech stack in their competitive environment and assume comparative indicators of the risk of errors and formulate a qualitative plan for further optimization of the tech stack to retain only effective technologies that correspond to the level of competencies of the company's specialists.

Conclusion

In summary, this study continues and expands the debate on the development of absorptive capacity, contributes to the understanding of the limitations of absorptive capacity of companies and offers a quantitative approach to audit the tech stack of company websites.

Based on the results obtained, it can be assumed that the proposed quantitative approach will primarily be applicable to those companies that have not yet identified their unique tech stack. The use of unique technologies can contribute to a greater number of errors, including in those companies that do not have sufficient resources, such as team competencies and prior knowledge. In turn, the use of popular technologies, on the contrary, will contribute to fewer errors.

Significant directions for future research include retesting the developed quantitative approach on a large sample, as well as qualitative study of unique and popular technologies. Moreover, based on the mixed results obtained in relation to two prominent banks, digital leaders, a promising direction is to develop a factor model that characterizes the key factors that influence the assimilation and correct use of various technologies in their own digital products. A promising direction for further research is also the development of combined approaches to auditing the tech stack, including both quantitative and qualitative research. The development of such an approach should focus on maximizing the strengths of each of the combined approaches and minimizing the weaknesses.

Of course, this study has a few limitations. Thus, only the banking industry was considered; in the future, the proposed quantitative approach can be tested in many industries and a comparative assessment can be made. Another limitation is the small sample, primarily due to limited access to technology stack analysis by many banks. As another limitation, the approach of dividing technologies only according to indicators of uniqueness, rarity and popularity can be highlighted; however, in the future, other classifiers can be used for evaluation. ■

Appendix.

Materials have been published that take into account the number of errors recorded by researchers

https://bosfera.ru/bo/dorabotat-mir-2
https://bosfera.ru/bo/na-potreby-publike-god-spustya
https://bosfera.ru/bo/rabota-nad-kreditkami
https://bosfera.ru/bo/vklad-v-druzhelyubie
https://bosfera.ru/bo/dorabotat-mir
https://bosfera.ru/bo/vybiraem-mir
https://bosfera.ru/bo/na-potreby-publike
https://bosfera.ru/bo/kreditnaya-karta-onlayn-trudnosti-i-luchshie-praktiki-bankov

References

1. Schumpeter J.A. (1934) *The theory of economic development.* London: Oxford University Press.
2. Möller K. (2010) Sense-making and agenda construction in emerging business networks – How to direct radical innovation. *Industrial Marketing Management*, vol. 39, no. 3, pp. 361–371. https://doi.org/10.1016/j.indmarman.2009.03.014
3. Dell'Anno D., del Giudice M. (2015) Absorptive and desorptive capacity of actors within university-industry relations: does technology transfer matter? *Journal of Innovation and Entrepreneurship*, vol. 4, no. 1, corpus ID: 256236239. https://doi.org/10.1186/s13731-015-0028-2
4. Cohen W.M., Levinthal D.A. (1990) Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, vol. 35, no. 1, p. 128–152. https://doi.org/10.2307/2393553
5. Zahra S.A., George G. (2002) Absorptive capacity: A review, reconceptualization, and extension. *The Academy of Management Review*, vol. 27, no. 2, pp. 185–203. https://doi.org/10.5465/amr.2002.6587995
6. Noblet J.-P., Simon E., Parent R. (2011) Absorptive capacity: a proposed operationalization. *Knowledge Management Research & Practice*, vol. 9, no. 4, pp. 367–377. https://doi.org/10.1057/kmrp.2011.26

7. Caiazza R., Phan P., Lehmann E., Etzkowitz H. (2021) An absorptive capacity-based systems view of Covid-19 in the small business economy. *International Entrepreneurship and Management Journal*, vol. 17, no. 3, pp. 1419–1439. https://doi.org/10.1007/s11365-021-00753-7
8. Grandinetti R. (2016) Absorptive capacity and knowledge management in small and medium enterprises. *Knowledge Management Research & Practice*, vol. 14, no. 2, pp. 159–168. https://doi.org/10.1057/kmrp.2016.2
9. Lehmann E.E., Menter M., Wirsching K. (2022) University spillovers, absorptive capacities, and firm performance. *Eurasian Business Review*, vol. 12, no. 1, pp. 125–150. https://doi.org/10.1007/s40821-021-00199-5
10. Arrighetti A., Vivarelli M. (1999) The role of innovation in the postentry performance of new small firms: Evidence from Italy. *Southern Economic Journal*, vol. 65, no. 4, pp. 927–939. https://doi.org/10.2307/1061285
11. Audretsch D.B. (1995) Firm profitability, growth, and innovation. *Review of Industrial Organization*, vol. 10, no. 5, pp. 579–588. https://doi.org/10.1007/bf01026883
12. Calvo J.L. (2006) Testing Gibrat's law for small, young and innovating firms. *Small Business Economics,* vol. 26, no. 2, pp. 117–123. https://doi.org/10.1007/s11187-004-2135-5
13. Colombelli A., Krafft J., Quatraro F. (2013) Properties of knowledge base and firm survival: Evidence from a sample of French manufacturing firms. *Technological Forecasting and Social Change,* vol. 80, no. 8, pp. 1469–1483. https://doi.org/10.1016/j.techfore.2013.03.003
14. Wagner S., Cockburn I. (2010) Patents and the survival of Internet-related IPOs. *Research Policy*, vol. 39, no. 2, pp. 214–228. https://doi.org/10.1016/j.respol.2009.12.003
15. Samuelsson M., Davidsson P. (2009) Does venture opportunity variation matter? Investigating systematic process differences between innovative and imitative new ventures. *Small Business Economics*, vol. 33, no. 2, pp. 229–255. https://doi.org/10.1007/s11187-007-9093-7
16. Scherer F.M., Harhoff D. (2000) Technology policy for a world of skew-distributed outcomes. *Research Policy*, vol. 29, nos. 4–5, pp. 559–566. https://doi.org/10.1016/s0048-7333(99)00089-x
17. Brown J.R., Martinsson G., Petersen B.C. (2012) Do financing constraints matter for R&D? *European Economic Review*, vol. 56, no. 8, pp. 1512–1529. https://doi.org/10.1016/j.euroecorev.2012.07.007
18. Minetti R. (2010) Informed finance and technological conservatism. *Review of Finance*, vol. 15, no. 3, pp. 633–692. https://doi.org/10.1093/rof/rfq024
19. Burford N., Shipilov A., Furr N. (2023) How GDPR changed European companies' tech stacks. *Harvard Business Review Digital Articles*. Available at: https://hbr.org/2023/02/how-gdpr-changed-european-companies-tech-stacks (accessed 10 December 2023).
20. Spiller C. (2022) Why and how planners need to audit their tech stack: Practice leaders need a clear view of their firm's tech capabilities and costs. *Journal of Financial Planning*, vol. 35, no. 12, pp. 60–63.
21. Weekes S. In focus: Creating a great recruitment tech stack. *Recruiter,* pp. 16–17.
22. Heap (2023) *What is a Tech Stack: Examples, components, and diagrams*. Available at: https://www.heap.io/topics/what-is-a-tech-stack (accessed 10 December 2023).
23. Tax W.K. (2021) Accounting techniques to keep your tech stack on the leading edge. *Accountingtoday.Com*, N.PAG.
24. Gvozdetskaya S. (2021) *Landing efficiency: How to audit a one-pager*. Available at: https://vc.ru/services/264089-effektivnost-lendinga-kak-provesti-reviziyu-odnostranichnika (accessed 10 December 2023) (in Russian).
25. Park H., Reber B.H. (2008) Relationship building and the use of Web sites: How Fortune 500 corporations use their Web sites to build relationships. *Public Relations Review*, vol. 34, no. 4, pp. 409–411. https://doi.org/10.1016/j.pubrev.2008.06.006

26. Digital transformation sets: starting conditions and priorities (2021) *April 22 intl. scientific conf. on Problems of Development of the Economy and Society, Moscow, 13–30 April 2021.* Moscow: HSE University, 2021 (in Russian).

27. Plusworld.ru (2023) *How sanctions affected the digitalization of banks*. Available at: https://plusworld.ru/articles/51319/ (accessed 10 December 2023) (in Russian).

28. Banking Review (2022) *Online applications for banking products: errors and opportunities*. Available at: https://bosfera.ru/bo/onlayn-zayavki-na-bankovskie-produkty-oshibki-i-i-vozmozhnosti (accessed 10 December 2023) (in Russian).

29. Semenikhin V.A., Mikaelyan A.M., Serova I.V., Tsilikov A.R. (2022) Isn't there too much technology? *Banking Review*, no. 7, pp. 68–71 (in Russian).

30. Laakso P. (2022) *Alfa-Bank topped the ranking of the best mobile banks for daily tasks on Android according to Markswebb*. Available at: https://vc.ru/finance/568392-alfa-bank-vozglavil-reyting-luchshih-mobilnyh-bankov-dlya-ezhednevnyh-zadach-na-android-po-versii-markswebb (accessed 10 December 2023) (in Russian).

31. Semenikhin V.A., Gladko A.M., Mikaelyan A.M. (2021) Speed and friendliness through the eyes of a new client. *Banking Review*, no. 9, pp. 64–69 (in Russian).

32. Tinkoff (2022) *Tinkoff Premium is recognized as the best digital bank for premium clients by Frank RG*. Available at: https://www.tinkoff.ru/about/news/02-11-2022-tinkoff-premium-best-digital-bank-for-preium-clients-according-to-frg/ (accessed 10 December 2023) (in Russian).

About the authors

Ivan V. Solovyov

Postgraduate Student, Lecturer, Department of Operational Management and Logistics, Graduate School of Business, HSE University, 26–28, Shabolovka Street, Moscow 119049, Russia;

E-mail: solovjev.iw@yandex.ru, isolovev@hse.ru

ORCID: 0000-0002-3120-9726

Viatcheslav A. Semenikhin

Postgraduate Student, Department of Operational Management and Logistics, Graduate School of Business, HSE University, 26–28, Shabolovka Street, Moscow 119049, Russia;

E-mail: vsemenikhin@planear.ru, vasemenikhin@hse.ru

ORCID: 0000-0001-6923-0363

Sergey P. Kushch

Dr. Sci. (Econ.), Professor;

Professor, Head of the Department of Operational Management and Logistics, Graduate School of Business, HSE University, 26–28, Shabolovka Street, Moscow 119049, Russia;

E-mail: skushch@hse.ru

ORCID: 0000-0002-6749-184X

DOI: 10.17323/2587-814X.2023.4.73.93

Methods and models for substantiating application scenarios for the digitalization of manufacturing and business processes of network enterprises

Yury F. Telnov
E-mail: telnov.yuf@rea.ru

Vasiliy A. Kazakov
E-mail: kazakov.va@rea.ru

Aleksey A. Bryzgalov
E-mail: bryzgalov.aa@rea.ru

Igor G. Fiodorov
E-mail: fedorov.ig@rea.ru

Plekhanov Russian University of Economics
Address: 36, Stremyanny Lane, Moscow 117997, Russia

Abstract

The process of digital transformation of enterprises is associated with the organization of manufacturing and business processes within the framework of selected types of business models and digital platforms, the distribution and economic substantiation of the roles of participants in network interactions, and ensuring the semantic interoperability of their interaction. Currently, certain experience has been accumulated in the implementation of modern business models for the digital transformation of enterprises which is reflected in the concepts of the Industrie 4.0, the Industrial Internet of Things, the creation of cyber-physical production systems, smart enterprises and intelligent manufacturing. At the same time, the issues of conceptual modeling of the architecture of digital enterprises, which determines the construction of manufacturing and business processes, and its economic substantiation depending on various factors of the external environment and internal economic potential have not yet been sufficiently researched and developed. All of the foregoing determines the relevance of the work presented here. The purpose of the study was to develop ontological and economic methods for substantiating application scenarios for the digitalization of manufacturing and business processes depending on the selected types of business models and digital platforms. To solve the problem, methods of classification, ontological engineering, activity-based costing and analysis of cash flows of income and expenses are used. The article presents an analysis of enterprise digitalization scenarios depending on the types of manufacturing and business processes, the types of business models and digital platforms used. An ontological model of enterprise digital transformation has been constructed, providing a choice of application scenarios for the digitalization of manufacturing and business processes for various types of business models and digital platforms. An economic model is proposed to justify options for constructing application scenarios for the digitalization of production and business processes depending on the distribution of roles of participants in network interaction using methods of activity-based costing and cash flow analysis.

Keywords: network enterprise, application scenario of digitalization, type of business model, type of digital platform, manufacturing and business processes, ontological model, economic model

Citation: Telnov Yu.F., Kazakov V.A., Bryzgalov A.A., Fiodorov I.G. (2023) Methods and models for substantiating application scenarios for the digitalization of manufacturing and business processes of network enterprises. *Business Informatics*, vol. 17, no. 4, pp. 73–93. DOI: 10.17323/2587-814X.2023.4.73.93

Introduction

The implementation of modern digital technologies in industry leads to a transition from mass production to mass customization, whereby the manufacturing of small-scale and individual products becomes predominant [1]. The basis of digital transformation in industry is the use of advanced production technologies based on the industrial internet of things (IIoT), cyber-physical systems (CPS), digital twins and platforms, and artificial intelligence. In this regard, the requirements for enterprise management systems are changing in terms of adaptability of operational management and flexibility in configuring production chains not only at the level of enterprise, but also at the level of interaction between enterprises within the formation of networked business structures (network enterprises).

At the present stage, the organization of flexible production is based on the use of cyber-physical systems, intelligent assets equipped with RFID and supported by Industrial Internet of Things technology, and cyber-physical production systems (CPPS) which combine individual assets into systems using digital twins and digital threads technology on the workshops, factories and supply chains levels [2–4].

The development of the Industry 4.0 modern production technologies significantly changes the architecture of enterprise management systems and the time intervals for operational and tactical planning and process regulation. First of all, the organization of digital production systems is characteristic of the level of management of individual equipment and production lines. At the same time, the spread of digital twin technology, implemented using multi-agent systems, to organizational entities that are participants in the value-added chain of networked enterprises, makes it possible to modernize manufacturing and business process management systems.

Existing digital twin systems, described in [5–7], are mainly focused on displaying the state of objects in the physical and/or virtual world with the organization of access to this information to all stakeholders involved in real manufacturing or business processes of the enterprise or managing these processes. Digital twins also allow for simulation of manufacturing and business processes to optimize the use of enterprise resources [8–10].

To improve the efficiency of production systems, the concept of an industrial agent is being actively developed. This is understood as “an autonomous and self-sufficient cyber-physical entity that represents the functionality of one or more industrial assets and manages them, providing permanent or temporary physical communication in order to perform functions and processes” [11]. The autonomy of industrial agents means their ability to respond to events arising in the external environment, making decisions and their execution. The construction of such production systems is based on the use of intelligent technologies associated with the development of dynamic multi-agent systems.

To integrate participants in network enterprises, the issues of creating and using digital platforms and connected intelligent agents come first. This should ensure the implementation of the principles of decentralization of management, vertical and horizontal integration of manufacturing and business processes, rapid reconfiguration of production chains and increased reliability of the entire production system [3, 4, 12].

From a technological point of view, the digital platform is a set of software services united by a common software environment to implement various functions of creating and operating a business ecosystem and individual network enterprises [13].

The work [14] summarizes the experience of using digital platforms at various industrial enterprises in Germany and Japan, thanks to which the authors propose a classification:

- The cloud platform implements, using cloud services, centralized collection and processing of manufacturing companies’ data, which is processed for the purpose of timely diagnosis of deviations from plan and long-term performance forecasting of the production structure. Open cloud platforms operate for the entire business ecosystem, while closed cloud platforms operate only for participants in individual digital or networked enterprises.
- The edge platform extends the cloud platform with computing infrastructure deployed locally at remote sites corresponding to production assets such as equipment, production lines, workshops and factories. In this regard, the collection and primary processing of data is carried out at remote sites in a closed mode, and summarizing information and making centralized decisions is possible in an open cloud environment.
- The brokerage platform (market place) takes on the functions of organizing the interaction of enterprises with each other in terms of selecting the best business partners according to various criteria, and plays the role, in the simplest case, of a trading platform. Intermediary (brokerage) platforms usually have an open nature of forming a business ecosystem.
- A hybrid platform allows us to combine the functionality of different types of digital platforms for different types of business models of networked enterprises.

The choice of the digital platform type is closely related to the choice of the business model type of the production system, which determines the pattern of interconnected material, information and financial flows from the perspective of the overall digital transformation strategy, taking into account technological and resource limitations [15]. As a rule, this choice is one-to-one, that is, the business model determines the requirements for the digital platform, and the digital platform sets restrictions on the implementation of the business model.

In [16], a generalized classification of business models for Industry 4.0 systems is given, according to which the following are distinguished: the industrial internet of things platform model, the value-adding services in operation model, the brokerage platform model, and the data trustee model. In [15], a multi-criteria model for choosing the type of business model was proposed, considering the network effects obtained, factors of the company's digital maturity, commercial risks and information security risks.

In principle, one enterprise can use different types of business models (BM) depending on the characteristics of the type of value-added chain and the life cycle stage of the products and/or services provided [17]. Consequently, the features of the enterprise's operating environment determine the requirements for building a business model, and the business model can radically transform the corresponding manufacturing and business processes using a certain application scenario of digitalization (use of digital technologies) [18]. Thus, types of digital platforms, types of business models and application scenarios of digitalization turn out to be highly interrelated, influencing each other.

At the same time, the implementation of modern digital technologies in the functioning of industrial enterprises remains an insufficiently researched area that requires generalization of the accumulated experience in the practical application of production technologies and the formation of a methodology for justifying the choice of certain application scenarios of digitalization in connection with the choice of types of business models and digital platforms.

The Plattform Industrie 4.0 AG2 (R&D Working Group) [19] generalized the practice of using digital platforms to organize manufacturing and business processes and proposed promising application scenarios for new projects of the digital transformation of enterprises. Based on the selected application scenarios, it is possible to build test benches on which it is possible to test various operating modes of enterprises. Similar work on the formation of standard scenarios of digitalization was carried out by the Smart Service Welt working group [20, 21] and in the approach to creating test benches of the Industrial Internet Consortium (IIC) [22], which allow for testing the proposed use cases for application scenarios, exploring promising technology development scenarios and forming standardization requirements.

A comparison of the listed approaches showed a very strong overlap in the content of the considered aspects of digitalization in the application scenario based on an analysis of the value of the collected operation data of assets, and the scenario for ensuring the transparency and adaptability of the supplied products. At the same time, the approach to building application scenarios in the concept of the Platform Industry 4.0 project is more complete in terms of implementing processes for all main types of manufacturing and business processes at various life cycle stages. Therefore, in the future, this approach will serve as the basis for studying application scenarios of digitalization for various types of business models and digital platforms.

The accumulation of experience in the implementation of business models, digital platforms and application scenarios of digitalization and its generalization in the form of reference models makes it possible to organize a knowledge-based system [23], which would make it possible to select appropriate scenarios and business models for the digital transformation of enterprises based on qualitative criteria, with the need to combine and adapt them to the operating conditions of a particular enterprise and calculate direct network effects from the application of selected scenarios. Moreover, all tasks are solved by constructing an ontology of digital transformation of enterprises, and the last task is based on the use of a combination of activity-based costing of performing manufacturing and business processes, and cash flow analysis to assess ROI in digital transformation. In accordance with the problem statement presented here, the article aims to develop methods and models for substantiating a scenario for the digitalization of manufacturing and business processes of enterprises, taking into account the choice of the type of business model and digital platform.

1. Methods for substantiating scenarios for digital transformation of an enterprise's manufacturing and business processes

From an architectural point of view, the transformation of enterprises based on digital technologies is carried out at several architectural levels [24, 25]:

- business organizations – identification of stakeholders, their vision of digital transformation, declared values, goals and objectives of the enterprise's digitalization;
- user participation – defining a sequence of activities involving users that provide the necessary functionality to achieve the capabilities of the digital production system;
- functional requirements – identifying the functional components of a digital production system, determining their structure and relationships, interfaces with the external environment;
- implementations – the use of technology to implement functional components, their communication patterns and life cycle procedures.

This article proposes methods for substantiating the digitalization of manufacturing and business processes of an enterprise which are determined by the interrelated choice of an application scenario of digitalization, the type of business models and the type of digital platform and it provides requirements for a digital production system at the level of business organization and user participation. The relationship between application scenarios of digitalization, types of business models and types of digital platforms is presented in the table of correspondence between the components of digital transformation of enterprises (*Table 1*), which is based on [14, 16, 19].

Examples of the implementation of application scenarios of digitalization based on the use of various types of business models and digital platforms in practice are reflected in works [26–31].

The proposed methodology for substantiating the digital transformation of enterprises is implemented within the framework of the created knowledge-based system and includes the consistent application of the following methods:

- Carrying out ontological engineering and analysis of the applicability of various application scenarios for the digitalization of manufacturing and business processes to the operating conditions of a particular enterprise as a result of which application scenarios of digitalization, types of business models and types of digital platforms that make up specific use cases are selected for various types of value chains.
- Perform an economic analysis of the applicability of selected use cases as a combination of application scenarios, business model type and digital platform type, based on the calculation of direct network effects for all parties involved.

When performing the stage of ontological engineering and analysis, it is proposed to reflect in the ontology of digital transformation of enterprises the types of processes within value-added chains and scenarios for their digitalization [19, 22, 32], types of business models [16, 17], types of digital platforms [14], factors for the need for digital transformation of manufacturing and business processes and factors for the choice of types of business models [15].

When describing application scenarios for digitalization of enterprises, it is necessary to define such main sections as [19, 22, 32]: stakeholders in digital transformation (actors); their roles in the transformation process; for each role, a vision of their implementation; key values and experience that the actor receives as a result of the implementation of the scenario; fundamental capabilities that characterize the features of ongoing innovations from the perspective of implemented technologies.

The basis for constructing an ontological representation of the type of business model and the type of digital platform is the framework for constructing a business model of St. Gallen [14–17, 33], which distinguishes the following main categories:

- participants in manufacturing and business processes, and their roles;
- value proposition at the output of processes, the result of the process;

Table 1.

Correspondence of components of digital transformation of enterprises

Value-added chain processes	Application scenario of digitalization	Business model type	Digital platform type
Product life-cycle management	IPD – value-added chain "Innovative Product Development," creating a concept and designing a product	Data trustee	Cloud platform
	SP2 – value-added chain "Smart Product Development for Smart Production," full cycle of development of intelligent products	Data trustee; IIoT platform model	Cloud platform
Production system life-cycle management	SPD – value-added chain "Seamless and dynamic plant engineering," organization and equipment of the factory (workshop)	Value adding services in operation	Edge platform
	AF – value-added chain "Adaptable Factory," management of production resources in the manufacturing process	Value adding services in operation	Edge platform
Supply chain management	OCP – value-added chain "Order-Controlled Production," managing the distribution of a common pool of resources between value chain participants	Value adding services in operation	Brokerage platform (marketplace)
			Cloud platform
	SAL – value-added chain "Self-organizing Adaptive Logistics," logistics routing	Value adding services in operation	Brokerage platform (marketplace)
			Cloud platform
Service	VBS – value-added chain "Value-Based Services"	IIoT platform model	Cloud platform
	TAP – value-added chain "Transparency and Adaptability of delivered Products," management and trusted access to product data	Data trustee	Cloud platform

- value chain, which determines the characteristics of the implementation of key work and the interaction of process participants;
- revenue mechanism that determines cash flows between process participants that create value for process consumers, as well as possible cost savings on work.

In addition, for digital platforms such additional characteristics are specified as [14]:

- features of concluding business contracts;
- description of key business model innovations (description of changes in the business model, qualitative characteristics of network effects);
- features of information security.

To carry out ontological analysis, the ontology of digital transformation of enterprises can be used in two modes:

- in reference mode, when any ontology category can be displayed to the decision maker for study with the necessary detail of properties and relationships;
- in the mode of selecting and justifying the use of certain components of digital transformation: appli-

cation scenarios, types of business models, types of digital platforms separately and in conjunction with each other from the perspective of various aspects and their combination when building value-added chains.

- In the second case, the query specifies the basic parameters of the enterprise, such as the type of enterprise, the type of production system, the characteristics of the product being manufactured, the life-cycle stage, and the proposed types of transformed manufacturing and business processes. Recommendations are provided in the corresponding responses to queries.

For the level of architecture of the digital production system, which characterizes the implementation of application scenarios of digitalization, the ontology of digital transformation of enterprises is expanded by describing the points of view on the use of the scenario from the position of each of the participants.

The description of the point of view covers a description of the type of activity associated with the application scenario use, which in turn includes the condition for its execution, the results obtained, restrictions and descriptions of the sequence of tasks (works) performed.

Based on such a detailed presentation of application scenarios, it becomes possible to carry out the second stage of the methodology for substantiating the enterprise's digital transformation, associated with an economic analysis of the possibility of implementing the scenario.

The essence of economic analysis comes down to assessing the network effect for each participant in network interaction according to the selected application scenario of enterprise digitalization. The effect of an individual participant is defined as the difference between the income received from the provision of services in the value chain and the costs associated with payments for the use of services provided by other participants in the value chain and the costs of performing the work themselves. The income and costs of each participant in the value chain are calculated on the basis of known payment items recorded in the "revenue mechanism" section of the ontological description of the type of business model (type of digital platform). The costs of performing their own work by each participant in the value chain are calculated using the method of activity-based costing for a set of tasks performed for each type of activity. In the same way, one-time costs for creating a digital platform and organizing manufacturing and business processes can be calculated. The method of economic analysis of the effectiveness of implementing an application scenario is discussed in detail in the corresponding section taking the example of two options for using a VBS application scenario.

2. Ontological model of digital enterprise transformation

Using the ontology of digital transformation of enterprises to create a knowledge-based system (KBS) that allows us to form business models and application scenarios for their use in specific digital enterprises provides the solution to the following tasks:

- Firstly, the ontology allows you to define a classification of typical business models, digital platforms and application scenarios according to which any digital enterprise can formulate appropriate requirements for its architecture.
- Secondly, in accordance with the ontology, consulting companies can accumulate knowledge bases of real precedents for using models of digital transformation of enterprises which can be selected by analogy and adapted to the operating conditions of specific digital enterprises.

The interaction scenario between the user and KBS is shown in *Fig. 1.*

One of the key elements of KBS is the digital transformation ontology based on the principles of constructing a design ontology [34] and reflecting methods and models for designing components of a digital transformation option (such as a business model, a type of digital platform and an application scenario of digitalization) of a networked enterprise focused on the implementation of the Industry 4.0 concept.

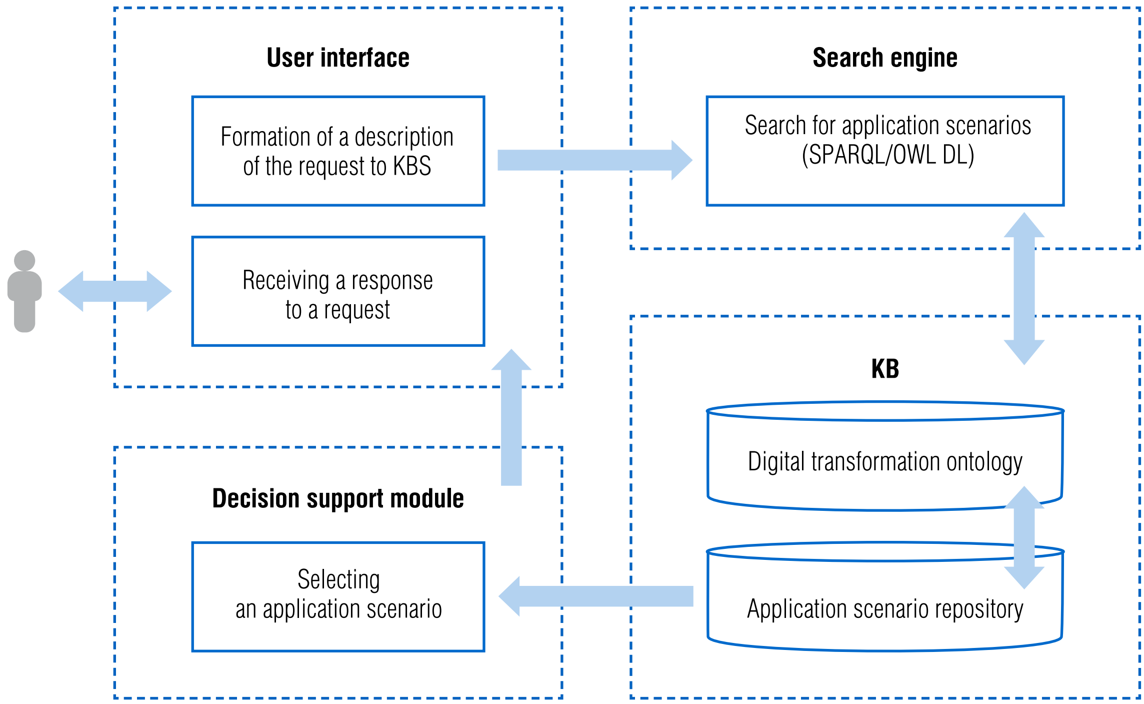


Fig. 1. Interaction between the user and the knowledge-based system to justify the option of digital transformation of the enterprise.

The user, through the interface, generates a query to KBS to select components of the digital transformation option and receives a response from KBS accordingly.

The search service selects components of the digital transformation option that correspond to the current problem situation defined by the user of KBS using the digital transformation ontology.

A knowledge-based system allows you to effectively accumulate and systematize the best experience of digital transformation projects in a repository that stores typical digital transformation options and specific implementations of application scenarios of digitalization.

The decision support module implements a qualitative and cost analysis of digital transformation options selected from the repository according to [35].

Let us consider the ontological model of digital transformation of a networked enterprise in more detail.

The key concepts of the digital transformation ontology are the traditional business entities of the business modeling ontology [36]: Enterprise, Product, Life-cycle Type and Life-cycle Stage, Strategy, Business Process, Business Model class, Roles, their Tasks, Risks, Costs, Value Propositions, and place in the Value Chain, the Revenue Mechanism as a whole. Along with this, for the purposes of KBS, the concepts of the domain of the Industry 4.0 are integrated into the ontology: digital platform, business model type, application scenario and its types according to [19].

The developed structure of the digital transformation ontology is divided conditionally for clarity into two parts, respectively related to the choice of the type of business model (*Fig. 2*) and the choice of an applied scenario of digitalization (*Fig. 3*).

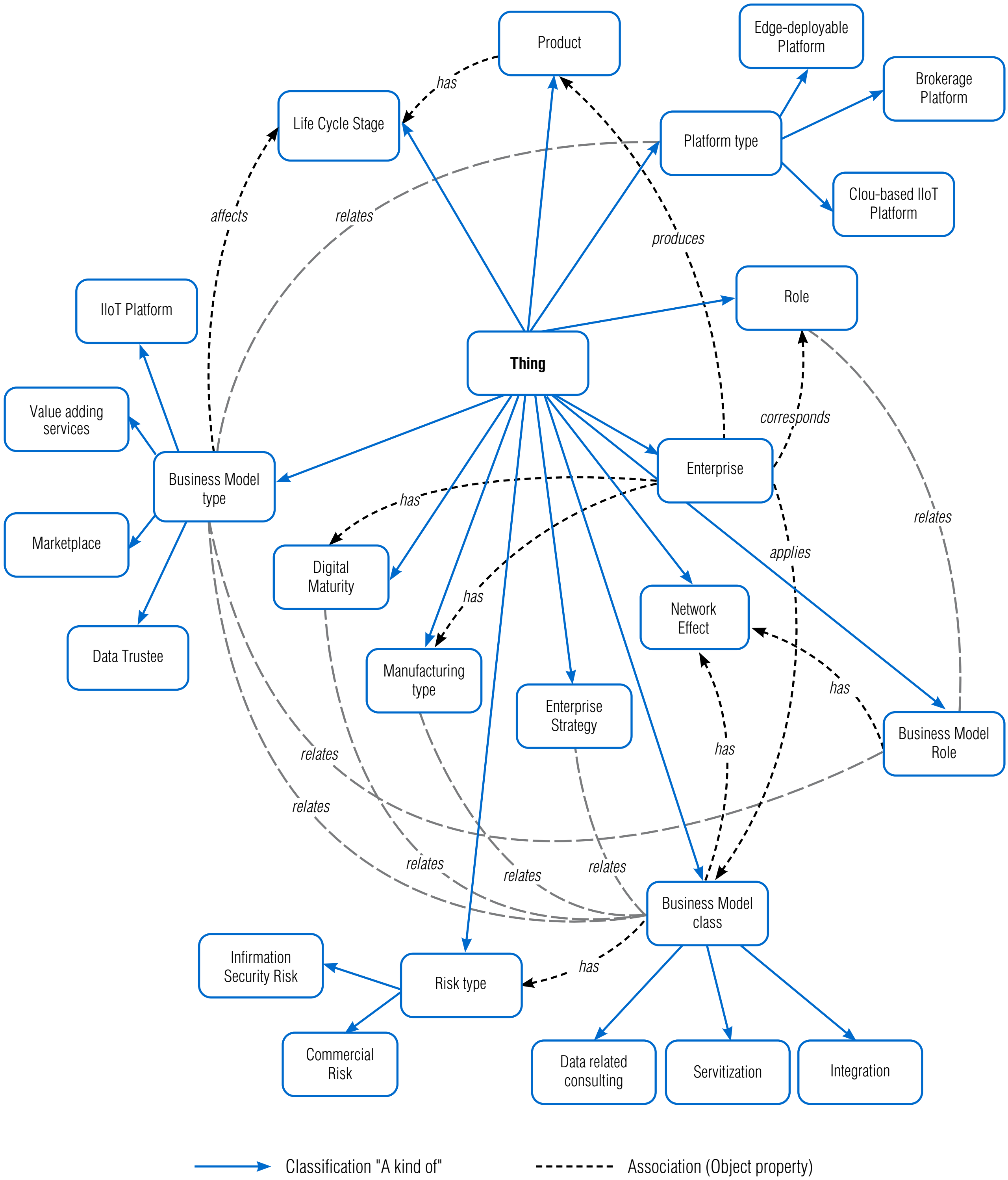


Fig. 2. Digital transformation ontology (part "Selecting a business model type").

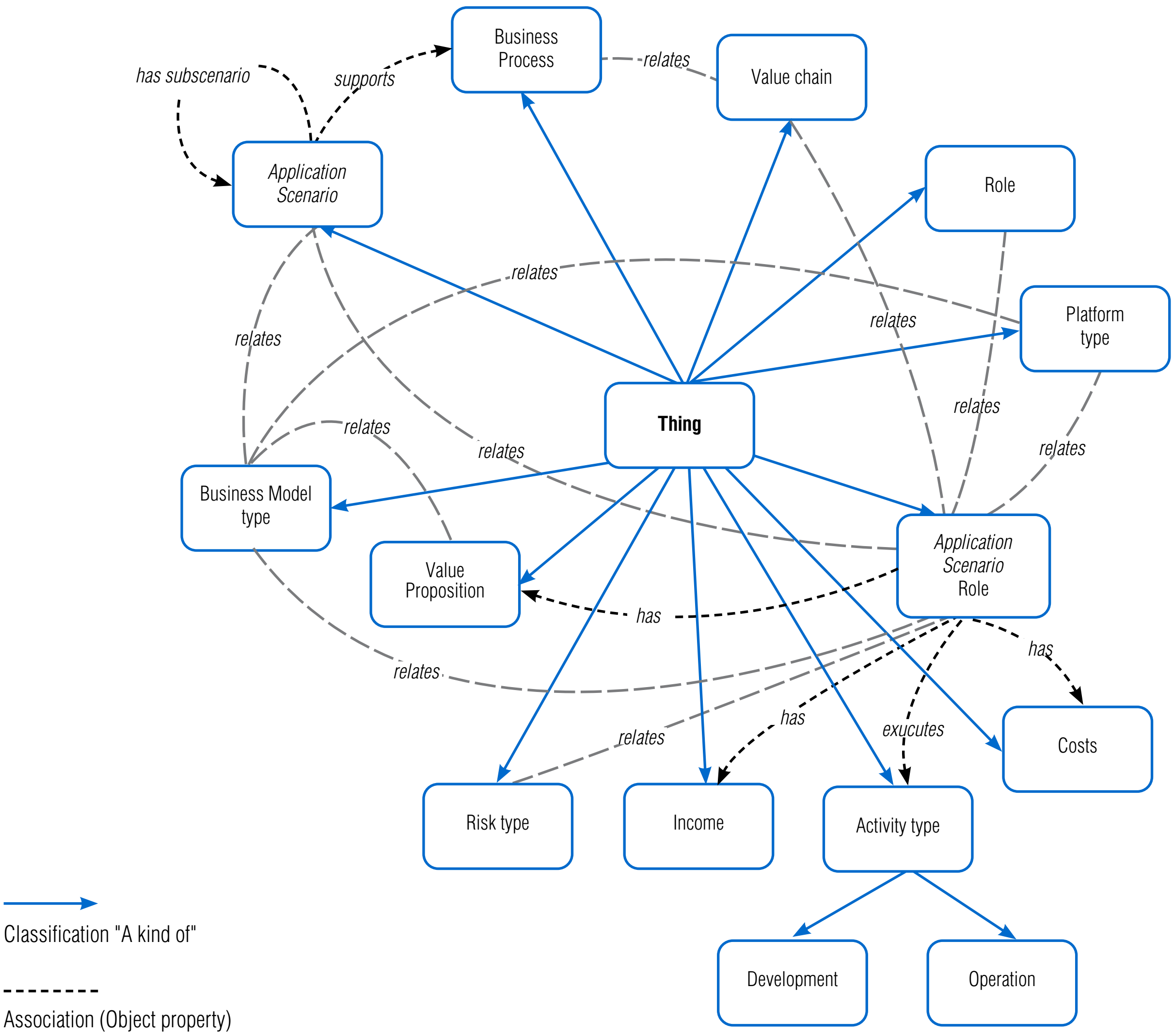


Fig. 3. Digital transformation ontology (part "Selecting an application scenario").

To justify the choice of the type of Industry 4.0 business model (type of digital platform), interrelated descriptions are specified that determine the correspondence of the characteristics inherent to the organization (corporate strategy, manufactured products and production system) and the characteristics of a possible business model (the required level of digital maturity, possible risks, network effects by digital transformation) (*Fig. 2*). Further actions of the KBS user allow you to select an option for the type of business model that meets the parameters of a specific organization [15].

The selected type of business model characterizes the way of organizing the functioning of a network enterprise based on a digital platform which can be detailed using typical application scenarios of digitalization [28, 30] (*Fig. 3*). The selection of application scenarios of digitalization tied to a specific type of business model and including a description of roles, income and cost structure, and a set of tasks to be performed is carried out based on the execution of standard queries to the KBS repository.

The main entity for describing an Application Scenario is an Application Scenario Role, the description of which includes a number of characteristic attributes:

- a typical Role performed within a specific Application scenario;
- the peculiarity of the Role's participation in the Value Chain;
- Value proposition offered by the Role;
- Value proposition related Income;
- investment and operational Costs borne by the participant in the networked enterprise performing the corresponding Role in the Application scenario (including those associated with receiving value from another participant);
- specific Platform Activity in the field of design and operation of the platform that is related to the activities of the corresponding Role;
- Commercial risks and Information security risks associated with the activities of the Role in the Application Scenario.

Typical requests are a reference to a description of application scenarios for using business models from an ontological knowledge base and the selection of specific application scenarios based on various characteristics that are of interest to stakeholders in the creation of networked enterprises.

With the further development of KBS, along with standard application scenarios, new application scenarios can be entered into the knowledge base, reflecting the best experience of the enterprise, including a problem situation description and the generated activity model. At the same time, concepts and instances are created that correspond to the description of the activity, problem situation and other elements of the application scenario. After this, relationships are established between the special application scenarios created and the existing standard application scenarios. In this way, the knowledge base can be regularly updated with up-to-date knowledge about effective ways to digitally transform an enterprise and organize value-added chains based on modern business models and digital platforms.

3. Economic model for substantiating the application scenario of digitalization of a networked enterprise

Justification of the feasibility of implementing various options for the structural organization of manufacturing and business processes formed on the basis of the type of business model, type of digital platform and application scenario of digitalization requires a quantitative economic analysis proving the possibility and effectiveness of their implementation. As a method for assessing the option of digital transformation of a networked enterprise, it is proposed to use the NPV (net present value) method, which allows you to link together all cash flows of different time periods and determine their total value at the current time [37, 38]. The purpose of applying the NPV method is to decide whether the parent enterprise and potential participants should invest in the organization of a networked enterprise.

Unlike a traditional enterprise, which independently invests in organizing its manufacturing and business processes, a networked enterprise requires minimal or no initial investment from all its participants. Investments are made from stakeholders depending on their goals in the value network and resource capabilities. This reduces the barrier to entry for many participants in the business ecosystem.

In general, to organize any application scenarios of digitalization in a networked enterprise, investments are required primarily in the creation of a digital platform, software services and agents interacting on the platform from the platform operator, service provider and service developer, respectively. The use of any implementation options for application scenarios generates a periodic cash flow for each participant in the value-added chain, which includes incoming and outgoing payments of the enterprise and its internal expenses.

The decision on the participation of a potential participant in the networked enterprise is made based on calculating the total NPV value for all its roles, which must be above a certain threshold value:

$$NPV_i = \sum_{j=1}^{r_i} NPV_{ij}, i = 1, ..., k, \qquad (1)$$

where NPV_i – net present value for i-th participant in the networked enterprise;

NPV_{ij} – net present value for j-th role of i-th participant in the networked enterprise

r_i – number of roles of i-th participant in the network enterprise;

k – number of the networked enterprise participants.

The main roles of participants in networked enterprises can be defined as follows [19]:

- Manufacturing Company;
- Equipment Supplier;
- Platform Operator;
- Service Provider;
- System Integrator;
- Service Developer;
- Platform Developer.

A networked enterprise usually includes a parent enterprise, most often with the role of a Manufacturing Company, which becomes the initiator of the project. It is engaged in the formation of orders for participants in the networked enterprise with the above roles. A networked enterprise, as a rule, is formed on the basis of a business ecosystem that has a digital platform.

The NPV value of one potential participant in the networked enterprise for each role in the value-added chain is determined by the formula:

$$NPV_j = -IC_j + \sum_{t=1}^{N} \frac{CF_{jt}}{(1-sd)^t}, j = 1, \ldots, r_i, \qquad (2)$$

where IC_j – initial investment for the j-th role of a participant in the networked enterprise;

CF_{jt} – cash flow of the t-th period (in a certain year) for the j-th role of a participant in the networked enterprise;

sd – discount rate;

sd – number of periods of existence of the network enterprise.

Cash flow is calculated for each year of existence of a networked enterprise as the difference between income from payments from other organizations and costs, including payments to other organizations, internal expenses and risk costs. It must be economically profitable for each potential participant to join a networked enterprise.

The assignment of roles to participants in a network enterprise is carried out in accordance with the application scenario of digitalization, for which its own characteristic set of value-added chain roles is determined depending on the focus of use, for example, on product management processes, on production system management processes, on supply chain management processes or on service processes. At the same time, as part of the implementation of an application scenario, an enterprise can perform several roles, or many enterprises can participate in one role.

In the future, we will consider the formalization of the economic model for justifying the application scenario of digitalization using the example of the applied scenario "value-based service" (VBS) [28]. The essence of the application scenario is that a Manufacturing Company rents machines from an equipment supplier that can operate using IIoT technology and, if necessary, be created to individual requirements. The Software Service Provider supplies software services to the platform which are operated by the Platform Operator. The Platform Operator receives data from the manufacturing enterprise machines connected to the digital platform in three ways and offers services to the parent enterprise based on the received data. The services consist of an analysis of the received data followed by the formation of recommendations from the Platform Operator or Software Service Provider. In addition, the data may be transferred to the Software Service Provider for the development of new or updating existing software services. In this implementation of the application scenario, each role corresponded to one enterprise (*Fig. 4*).

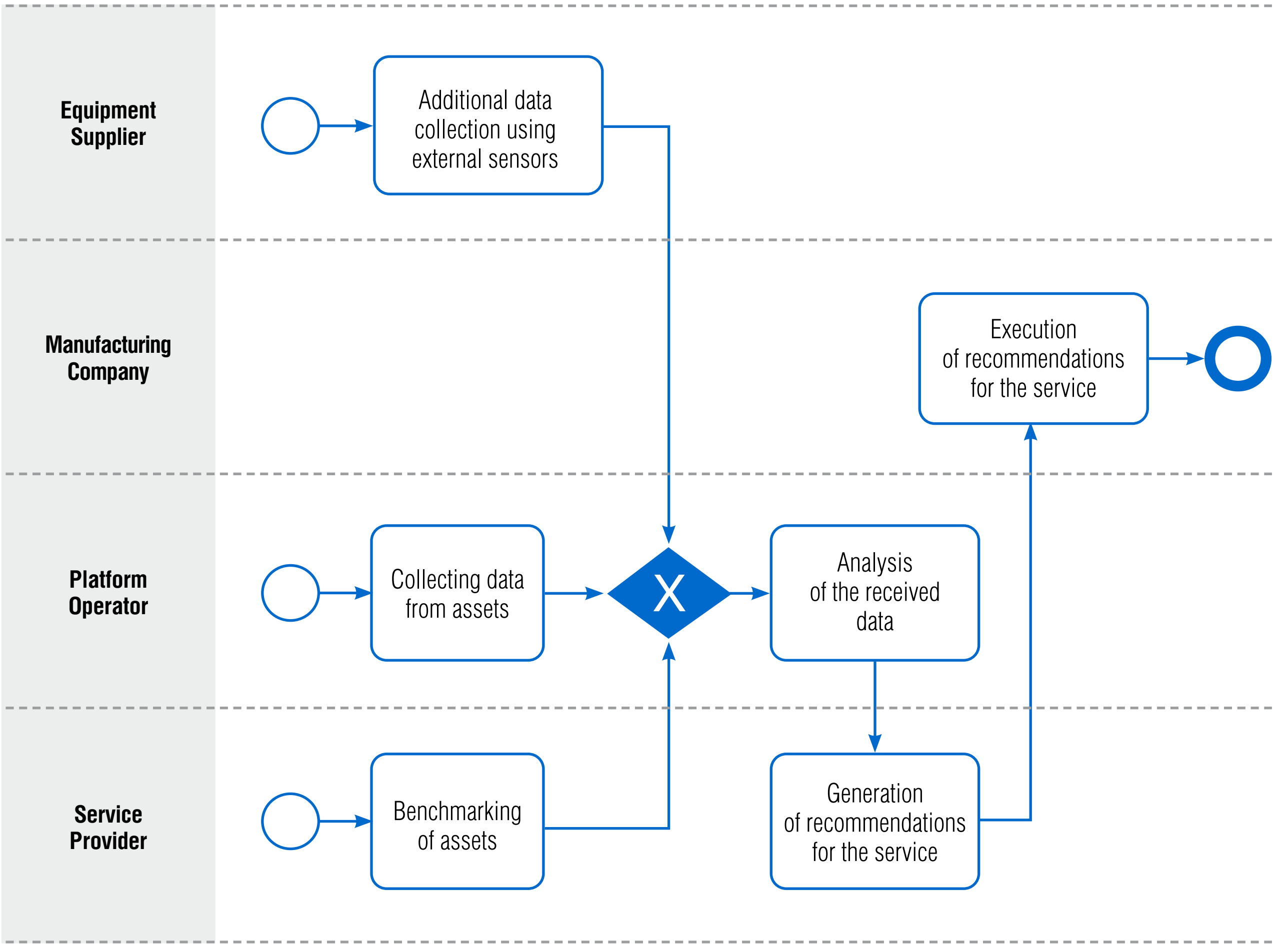


Fig. 4. The process of performing a service for collecting and analyzing of equipment operation data.

In another implementation of the VBS application scenario, the parent enterprise organizes its own digital platform, which allows it to control access to platform and software services. In this case, the parent company refuses to outsource work with the platform and assumes the role of Manufacturing Company and Platform Operator.

The decision to outsource digital services may be driven by the following factors:

- insufficient own financial and/or material resources to carry out certain business processes;
- insufficient motivation of the enterprise, for example, some business processes are not a priority for the enterprise, and their implementation can lead to a decrease in the rate of productivity of the main processes;
- the presence of restrictions, for example, related to confidential information, the leakage of which will lead to loss of competitiveness in the market.

In accordance with [28], the application scenario identifies work related to the organization of the platform and network enterprise which require initial investments from the following main roles of participants in the network enterprise: Platform Operator (PO), Service Provider (SP), Equipment Supplier (ES), Manufacturing Company (MC). The list of works and their distribution by network interaction participants is presented in *Table 2.*

The calculation of the total cost of initial investment (IC_i) for each role of the VBS application scenario can be represented as the summation of all types of costs under a specific role:

$$IC_j = \sum_{i=1}^{N} a_{ij}, i = 1, ..., k, \quad (3)$$

where a_{ij} – table element of i-th type of initial investment and j-th role of the application scenario;

N – number of cost types;

k – number of application scenario roles.

The use of application scenarios of digitalization leads to a change in the value-added chain and, accordingly, a new structure of costs and income for each of its participants. This is because there is a distribution of risks in the activities of a networked enterprise, fixed and variable costs between participants, as well as the emergence of new items of income and costs which depend on the performed roles of the participant in the value-added chain and the application scenario itself.

Thus, the changed structure of costs and income of the annual cash flow (CF_i) for each participant in the networked enterprise can be determined by the formula:

$$CF_i = DP_i + DN_i - SP_i - VZ_i, i = 1, ..., k, \quad (4)$$

where k – number of enterprises participating in the application scenario of digitalization;

DP_i – income from receiving payment for the services of i-th enterprise;

SP_i – own payments for services provided by i-th enterprise;

DN_i – indirect network effect of i-th enterprise;

VZ – internal costs of i-th enterprise, which are calculated based on the application of the functional cost analysis method [39–41].

For a manufacturing company, the indirect effect is a reduction in costs for equipment repair and main-

Table 2.

Initial investments

Investments in installation and configuration of equipment	Roles			
	PO	SP	ES	MC
Setting up the platform for the networked enterprise	IDP			
Service development		IDS		
Equipment development (custom requirements)			IDE	
Equipment rent				IE
Product equipment manufacturing			IPE	
Connecting and setting up equipment	ICN_1		ICN_2	ICN_3

tenance. For a software services provider, this means creating new services based on equipment operation data of a manufacturing enterprise which will lead to an increase in licenses. For the platform operator, this is a reduction in the costs of operating and supporting the platform due to an increase in the number of platform participants by increasing the attractiveness of the platform for networked enterprise participants through new software services and platform services. For the equipment supplier, this means reducing costs by reducing equipment downtime, since an increase in the number of production plants will make it more likely that the equipment will be rented.

In addition to the costs listed in *Table 3*, internal costs include fees for risk prevention activities. These fees may include the costs of analyzing and forecasting risks, the costs of eliminating the consequences of unforeseen risks and the costs of insurance against selected risks.

For a manufacturing company in a networked enterprise, the following risks associated with operational failures can be identified:

- quipment;
- platforms (including restricting access to them);
- software services (including restriction of access to them);

Table 3.

Cash flow for the year

Costs type (items of income/costs)	Roles			
	PO	SP	ES	MC
Service fee	S			$-S$
Payment for platform services from the service provider	SP	$-SP$		
Payment for platform services from the equipment supplier	SE	$-SE$		
Equipment connection fee	$-CN$		CN	
Equipment setup fee	$-F$		F	
Payment of a license for the use of the application by the equipment operator	$-L$	L		
Platform maintenance costs	$-PL$			
Costs of developing an application for the service		$-DA$		
Service development costs		$-DS$		
Equipment rental fee			AE	$-AE$
Payment for risk prevention activities	$-R_1$	$-R_2$	$-R_3$	$-R_4$
Indirect network effect	DN_1	DN_2	DN_3	DN_4

- failures in the supply of equipment for individual orders;
- incorrectly executed individual orders for equipment;
- lack of need for the product among customers.

For a software service provider in a networked enterprise, the following risks can be identified:

- loss of the ability to deliver software services to the platform;
- refusal of the platform operator to provide a software service due to lack of demand or the creation of a similar and competitive software service;
- impossibility of integrating custom-made equipment and software services;
- identifying critical errors in the software service;
- leak of confidential data about processes and users of a software service.

For the platform operator in a networked enterprise, the following risks can be identified:

- refusal to work on the platform of the parent enterprise;
- inability to operate the platform due to lack of financial resources or technological problems;
- loss of reputation due to low-quality software services;
- leak of confidential data about processes and users of platform services.

For the equipment supplier in a networked enterprise, the following risks can be identified:

- refusal to individually develop equipment;
- impossibility of integrating custom-made equipment and software services;
- refusal of a manufacturing company to rent equipment due to lack of demand or the creation of similar and competitive standard equipment.

In *Table 3*, the minus sign means payment for the service or internal costs; the absence of a sign means receipt of payment. The calculation of the cash flow value () for each role of the VBS application scenario can be represented as the summation of all types of costs under a specific role:

$$CF_j = \sum_{i=1}^{N} a_{ij}, i = 1, ..., k, \quad (5)$$

where a_{ij} – table element of i-th costs type and j-th role of the application scenario of digitalization;

N – number of costs types;

k – number of application scenario roles.

Given that an enterprise can perform several roles when organizing a networked enterprise, the number of types of incoming and outgoing payments changes. Depending on what roles enterprises take on within the application scenario of digitalization, the structure of risk accounting changes, as well as the main types of operating costs in a networked enterprise, when payments are replaced by internal costs.

The proposed economic model to justify scenarios for organizing manufacturing and business processes of a network enterprise based on the NPV method and activity-based costing allows us to assess the attractiveness of a network enterprise for all its potential participants. Thanks to economic analysis, deciding on the implementation of application scenarios of digitalization to create a network enterprise becomes economically justified due to the provision of information about possible income, costs, risks and other factors associated with the value-added chain. The analysis allows us to evaluate for each enterprise and its corresponding roles the potential profit generated because of payments, direct and indirect network effects and internal cost savings, compare them with initial costs, and also determine the best option for implementing the application scenario of digitalization.

Conclusion

An analysis of the experience of implementing business models, digital platforms and application scenarios of digitalization at enterprises shows the need to develop ontological and economic methods for the formation and justification of the organization of manufacturing and business processes depending on the type and potential of participants in networked enterprises. Moreover, the ontological model of digital transformation should serve as the basis for the formation of variants of application scenarios of digitalization for their subsequent economic justification.

The proposed method of ontological engineering and analysis of the feasibility of various application scenarios for the digitalization of manufacturing and business processes to the operating conditions of a particular enterprise involves mapping in the ontology the classification of types of business models, digital platforms and application scenarios themselves in interrelation for various types of value-added chains.

The article identifies the main types of queries to substantiate application scenarios for the digitalization of enterprises, which make it possible to select typical scenarios of digitalization based on individual or combinations of features that characterize the formation of value propositions, obtaining competitive advantages and ensuring positive cash flows depending on the performed roles of participants in network enterprises. The computer implementation of the ontology in the OWL format also allows for reference queries on the implementation of certain types of application scenarios.

In the future, the ontology of digital transformation of enterprises we developed can serve as the basis for accumulating an ontological database of precedents for the implementation of application scenarios, business models and digital platforms to search for the best practices of digital transformation and its adaptation to specific conditions.

Based on the economic analysis of network effects from the use of one or another option to construct an application scenario for the digitalization of enterprises, this article proposes the use of the NPV method, which determines investments in standard work preparing a digital platform for operation, as well as current income and costs in the form of mutual payments participants of network enterprises, considering the cost of performing internal work. From this point of view, the article defines the composition of income and cost items for a common application scenario to obtain value from the analysis of digital data. Comparison of the total network effect for various options for the role participation of stakeholders of a network enterprise allows you to select the best implementation of the application scenario. In the future, it is proposed to expand the method of economic analysis of options for constructing application scenarios by formalizing models for obtaining indirect network effects by expanding the number of participants in the business ecosystem.

The novelty of the proposed methods and models to justify options for the digital transformation of manufacturing and business processes of enterprises lies in the formulation and solution of the problem of the interrelated choice of the type of business model, type of digital platform and application scenario depending on the nature of the enterprise. At the same time, the ontological model of digital transformation serves as the basis for the formation of applied scenarios for the digitalization of enterprises, the choice of which is clarified because of applying the model for calculating economic efficiency using the NPV and activity-based costing methods. ■

Acknowledgments

This research was carried out thanks to a grant from the Russian Science Foundation № 22-11-00282, https://rscf.ru/project/22-11-00282/

References

1. Veselovsky M.Ya., Khoroshavina N.S. (eds.) (2021) *Digital transformation of industrial enterprises in an innovative economy.* Moscow: Mir Nauki. Available at: https://izd-mn.com/PDF/06MNNPM21.pdf (accessed 30 October 2023) (in Russian).

2. Makarov V.L., Bakhtizin A.R., Beklaryan G.L. (2019) Developing digital twins for production enterprises. *Business Informatics*, vol. 13, no. 4, pp. 7–16. https://doi.org/10.17323/1998-0663.2019.4.7.16

3. Salazar L.A.C., Ryashentseva D., Lüder A., Vogel-Heuser B. (2019) Cyber-physical production systems architecture based on multi-agent's design pattern – comparison of selected approaches mapping four agent patterns. *International Journal of Advanced Manufacturing Technology*, vol. 105, pp. 4005–4034. https://doi.org/10.1007/s00170-019-03800-4

4. Salazar L.A.C., Vogel-Heuser B. (2022) A CPPS-architecture and workflow for bringing agent-based technologies as a form of artificial intelligence into practice. *Automatisierungstechnik*, vol. 70, no. 6, pp. 580–598.

5. Technet (2019) *Digital twins in the high-tech industry*. Available at: http://assets.fea.ru/uploads/fea/news/2019/12_december/28/cifrovoy_dvoinik.pdf (accessed 30 October 2023) (in Russian).

6. Minaev V.A., Mazin A.V., Zdiruk K.B., Kulikov L.S. (2019) Digital twins of objects in the solution of control problems. *Radio Industry (Russia)*, vol. 29, no. 3, pp. 68–78 (in Russian). https://doi.org/10.21778/2413-9599-2019-29-3-68-78

7. Abramov V.I., Boboev D.S., Gilmanov T.D., Semenkov K.Yu. (2022) Theoretical and practical aspects of creating a company's digital twin. *Russian Journal of Innovation Economics*, vol. 12, no. 2, pp. 967–980 (in Russian). https://doi.org/10.18334/vinec.12.2.114890

8. Akopov A.S. (2012) Designing of integrated system-dynamics models for an oil company. *International Journal of Computer Applications in Technology*, vol. 45, no. 4, pp. 220–230. https://doi.org/10.1504/IJCAT.2012.051122

9. Akopov A.S. (2014) Parallel genetic algorithm with fading selection. *International Journal of Computer Applications in Technology*, vol. 49, no. 3/4, pp. 325–331. https://doi.org/10.1504/IJCAT.2014.062368

10. Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S. (2021) Digital plant: methods of discrete-event modeling and optimization of production characteristics. *Business Informatics*, vol. 15, no. 2, pp. 7–20. https://doi.org/10.17323/2587-814X.2021.2.7.20

11. Kovalenko I., Tilbury D., Barton K. (2019) The model-based product agent: A control oriented architecture for intelligent products in multi-agent manufacturing systems. *Control Engineering Practice*, vol. 86, pp. 105–117.

12. Komesker S., Motsch W., Popper J., Sidorenko S., Wagner A., Ruskowski M. (2022) Enabling a multi-agent system for resilient production flow in modular production systems. Proceedings of the *55th CIRP Conference on Manufacturing Systems, Procedia CIRP*, vol. 107, pp. 991–998.

13. Telnov Yu.F., Kazakov V.A., Danilov A.V., Denisov A.A. (2022) Requirements for the software implementation of the Industrie 4.0 system for creating network enterprises. *Software and Systems*, vol. 35, no. 4, pp. 557–571 (in Russian). https://doi.org/10.15827/0236-235X.140.557-571

14. Federal Ministry for Economic Affairs and Energy (BMWi) (2021) *Digital platforms in manufacturing industries. Result Paper*. Available at: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/digital-platforms-in-manufacturing-2021.pdf (accessed 30 October 2023).

15. Telnov Yu.F., Bryzgalov A.A., Kozyrev P.A., Koroleva D.S. (2022) Choosing the type of business model to implement the digital transformation strategy of a network enterprise. *Business Informatics*, vol. 16, no. 4, pp. 50–67. https://doi.org/10.17323/2587-814X.2022.4.50.67

16. Federal Ministry for Economic Affairs and Energy (BMWi) (2019) *Digital business models for Industrie 4.0. Result Paper*. Available at: https://www.bmwk.de/Redaktion/EN/Publikationen/Industry/digital-business-models-industry-4-0.pdf (accessed 30 October 2023).

17. Weking J., Stocker M., Kowalkiewicz M., Bohm M., Krcmar H. (2018) Archetypes for Industry 4.0 business model innovations. Proceedings of the *24th Americas Conference on Information Systems (AMCIS). Association for Information Systems (AIS)* (eds. A. Bush, V. Grover, S. Schiller).

18. Telnov Yu.F., Bryzgalov A.A., Koroleva D.S. (2023) Organization of production and business processes in value chains based on applied scenarios for the digitalization of enterprises. *Open Education*, vol. 27, no. 3, pp. 43–54 (in Russian).

19. Federal Ministry for Economic Affairs and Energy (BMWi) (2016) *Aspects of the research roadmap in application scenarios. Working paper.* Available at: http://www.plattform-i40.de/I40/Redaktion/EN/Downloads/Publikation/aspects-of-the-research-roadmap.html (accessed 30 October 2023).

20. acatech – Deutsche Akademie der Technikwissenschaften (2016) *Smart service WELT. Digitale Serviceplattformen – Praxiserfahrungen aus der Industrie. Best Practices*. Available at: https://innosabi.com/wp-content/uploads/2016/05/BerichtSmartService2016_DE_barrierefrei.pdf (accessed 30 October 2023).

21. Lu Y., Xu X., Wang L. (2020) Smart manufacturing process and system automation – A critical review of the standards and envisioned scenarios. *Journal of Manufacturing Systems*, vol. 56, pp. 312–325.

22. Federal Ministry for Economic Affairs and Energy (BMWi) (2016) *Proposal for a joint "scenario" of Plattform Industrie 4.0 and IIC. Discussion paper.* https://doi.org/10.13140/RG.2.2.33286.98881

23. Telnov Yu.F. (2005) *Reengineering of business-processes. Component-based methodology.* Moscow: Finance and Statistics (in Russian).

24. Federal Ministry for Economic Affairs and Energy (BMWi) (2019) *Usage view of asset administration shell. Discussion Paper.* Available at: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/2019-usage-view-asset-administration-shell.pdf (accessed 30 October 2023).

25. Industry IoT Consortium (2022) *The industrial internet reference architecture*. Available at: https://www.iiconsortium.org/wp-content/uploads/sites/2/2022/11/IIRA-v1.10.pdf (accessed 30 October 2023).

26. Seitz M., Gehlhoff F., Salazar L.A.C., Fay A., Vogel-Heuser B. (2021) Automation platform independent multi-agent system for robust networks of production resources in Industry 4.0. *Journal of Intelligent Manufacturing*, vol. 32, pp. 2023–2041. https://doi.org/10.1007/s10845-021-01759-2

27. Savastano M., Amendola C., D'Ascenzo F. (2018) How digital transformation is reshaping the manufacturing industry value chain: The new digital manufacturing ecosystem applied to a case study from the food industry. *Network, Smart and Open. Lecture Notes in Information Systems and Organisation,* vol. 24, pp. 127–142. https://doi.org/10.1007/978-3-319-62636-9_9

28. Federal Ministry for Economic Affairs and Energy (BMWi) (2018) *Usage viewpoint of application scenario value-based service. Discussion Paper*. Available at: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/hm-2018-usage-viewpoint.pdf (accessed 30 October 2023).

29. Federal Ministry for Economic Affairs and Energy (BMWi) (2020) *Usage view "Seamless and dynamic engineering of plants." Discussion Paper*. Available at: https://www.bmwk.de/Redaktion/EN/Publikationen/Industry/usage-view-seamless-and-dynamic-engineering-of-plants.pdf (accessed 30 October 2023).

30. Federal Ministry for Economic Affairs and Energy (BMWi) (2017) *Benefits of application scenario value-based service. Working paper*. Available at: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/benefits-application-scenario.html (accessed 30 October 2023).

31. Federal Ministry for Economic Affairs and Energy (BMWi) (2017) *Exemplification of the Industrie 4.0 application scenario value-based service following IIRA structure. Working Paper*. Available at: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/exemplification-i40-value-based-service.html (accessed 30 October 2023).

32. Federal Ministry for Economic Affairs and Energy (BMWi) (2019) *Guidance "Use cases and applications."* Available at: https://www.researchgate.net/publication/333719195_Guidance_Use_Cases_and_Applications (accessed 30 October 2023).

33. Gassmann O., Csik M., Frankenberger K. (2014) *The business model navigator: 55 Models that will revolutionise your business*. Pearson Education Limited.

34. Borgest N.M. (2023) Design ontology: Genesis and development. Proceedings of the *Twenty-first National Conference on Artificial Intelligence with international participation KII-2023 (October 16–20, 2023)*, pp. 6–13.

35. Bryzgalov A.A., Telnov Yu.F. (2022) An economic model for creating a network enterprise architecture. *Statistics and Economics*, vol. 19, no. 6, pp. 53–62 (in Russian). https://doi.org/10.21686/2500-3925-2022-6-53-62

36. Osterwalder A., Pigneur Y., Tucci C. (2010) Clarifying business models: Origins, present, and future of the concept. *Communications of the Association for Information Systems*, vol. 16. https://doi.org/10.17705/1CAIS.01601

37. Margolin A.M. (2018) *Economic assessment of investment projects*. Moscow: Economics (in Russian).

38. Vorotnikova D.V. (2022) Comparative analysis of the real options method and traditional methods for assessing the effectiveness of investment projects. *Bulletin of the Altai Academy of Economics and Law*, no. 2–1, pp. 24–27 (in Russian). https://doi.org/10.17513/vaael.2053

39. Drury C. (2007) *Management and cost accounting*. Cengage.

40. Yang C.H., Lee K.C., Li S.E. (2020) A mixed activity-based costing and resource constraint optimal decision model for IoT-oriented intelligent building management system portfolios. *Sustainable Cities and Society*, vol. 60, article 102142. https://doi.org/10.1016/j.scs.2020.102142

41. Gholami H., Jiran N.S., Saman M.Z.M., et al. (2019) Application of activity-based costing in estimating the costs of manufacturing process. *Transformations in Business and Economics*, vol. 18, no. 2, pp. 839–860.

About the authors

Yury F. Telnov

Dr. Sci. (Econ.), Professor;

Head of the Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: telnov.yuf@rea.ru

ORCID: 0000-0002-2983-8232

Vasiliy A. Kazakov

Cand. Sci. (Econ.);

Associate Professor, Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: kazakov.va@rea.ru

ORCID: 0000-0001-8939-2087

Aleksey A. Bryzgalov

Assistant Professor, Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: bryzgalov.aa@rea.ru

ORCID: 0000-0002-5001-4326

Igor G. Fiodorov

Dr. Sci. (Econ.);

Professor, Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: fedorov.ig@rea.ru

ORCID: 0000-0003-2335-0452

DOI: 10.17323/2587-814X.2023.4.94.112

A bibliometric review of scientific research on the significance of information technology relating to sustainable development reporting practice

Maneesh Kumar Pandey [a]
E-mail: maneeshban@gmail.com

Amit Kumar Pathak [b]
E-mail: Amit.Pathak@utas.edu.com

Irina G. Sergeeva [a]
E-mail: igsergeeva@gmail.com

[a] ITMO University
Address: 9, Lomonosov Street, St. Petersburg 191002, Russia

[b] University of Technology and Applied Sciences – Muscat
Address: PO Box 74, Al-Khuwair, Muscat 133, Sultanate of Oman

Abstract

Sustainable development, a prominent issue in the twenty-first century, is significantly influenced by the rapid global IT revolution. This study employs bibliometric analysis to explore the role of scientific research in sustainable development reporting, aligning with international standards and utilizing IT tools. It assesses countries' awareness of sustainable development reporting's importance in achieving socio-economic and environmental goals. The study examines article frequency, source countries, authors, co-authorship, citations, key term co-occurrences, and bibliometric coupling. The result concludes that active engagement among research work of academic institutions, government organizations, and industries of emerging countries on the development and role of information technology in sustainable development reporting practices can foster cost-effective ways for sustainable development reporting which may play a vital and crucial role in sustainable development reporting for middle- and low-income countries to ensure a green and sustainable future. This work can benefit middle- and low-income nations in their pursuit of a green and sustainable future. The research highlights the significance of academic institutions in enhancing sustainable development reporting, especially for Micro, Small, and Medium-Sized Enterprises (MSMEs) in middle and low-income countries, offering valuable insights for future actions, which in turn may help these countries to put more effort into this domain through their academic establishments.

Keywords: information technology (IT), sustainable development reporting (SDR), Scopus database, VOS viewer, bibliometric analysis

Citation: Pandey M.K., Pathak A.K., Sergeeva I.G. (2023) A bibliometric review of scientific research on the significance of information technology relating to sustainable development reporting practice. *Business Informatics*, vol. 17, no. 4, pp. 94–112. DOI: 10.17323/2587-814X.2023.4.94.112

Introduction

In the era of the 21st century, environmental preservation and sustainable development are no longer just topics of expert discussion or current fashion. The environmental, social and governance (ESG) sustainable development specifications are rapidly evolving, driven by increasing demands from investors, regulators, and other stakeholders. The need for high-quality disclosures and better use of data and key performance indicators (KPIs) is expected to be critical for a company's future. The imperative to build resiliency and sustainability has never been stronger; from the global push to reduce carbon emissions, the real impact of climate change on our lives, to the availability of funds to enable companies to meet their obligations and the technological advances that now support the ability to meet their ESG goals.

According to the UN's [1], sustainable development is one that "satisfies the demands of our time without jeopardizing the chances of coming generations from getting their resource requirements in future." Since then, the discussion of sustainable development and corporate behavior has gained momentum, with demands for innovative performance techniques, environmentally friendly company models and integrated reporting systems [2–4]. The Sustainable Development Goals (SDGs) advocated by the United Nations (UN), which establish the world's goals and aspirations for 2030, are encouraged by responsible corporate behavior, according to the 2030 Agenda.

Corporations play a crucial part in this process and may hold the key to success or the reason why this challenge fails [5]. Therefore, sustainability reporting has gotten a lot of attention from corporate management and national leadership during the last couple of years and numerous theoretical and empirical studies have been focusing on the sustainability reporting area. Companies are moving to go beyond financial performance in order to drive business because of the shifting global environment. Company executives are becoming more aware of the need to incorporate environmental and social issues into their overall company strategy. Hence, the SDGs are being included into organisations' thinking and reporting using integrated sustainability reporting, which has been described as a promising strategy for disclosing the business journey towards SDG achievement [6]. Therefore, large, multinational corporations and international organizations now routinely report on sustainability. Leading standards in the industry, such as the Global Reporting Initiative (GRI), Science-Based Targets Initiative for Net Zero or TCFD for disclosures, have made companies report progress towards goals with underlying information harvested from various data points across their value chain.

After extensive debate over the nature, constraints and challenges of integrating reporting, it has come to be widely acknowledged as one of the most important management and accounting developments across the corporate sector and other organizations, and it is getting wide acceptance across the globe [7, 8]. The International Integrated Reporting Council (IIRC) introduced integrated reporting in 2011 with the goal of providing a succinct and comprehensive report on how an organization's strategy, governance, performance, and prospects contribute to the creation of sustainable value over the short, medium, and long term.

For the purpose of creating non-financial reports, the Global Reporting Initiative (GRI) Guidelines and the Institute of Social and Ethical Accountability Standard (Accountability 1000, AA1000) have both been formed. Several organizations and businesses are using the widely accepted and implemented Guidelines for Sustainability Reporting standard established by the Global Reporting Initiative for presenting non-financial reports. However, in many low- and middle-income nations in Asia, Africa, and Latin America, there are a handful of firms or organizations that generate social and environmental reports. These non-financial reports increasingly need to be prepared by even medium- and small-sized firms due to the increased demand for them by national governments and other international bodies for monitoring sustainable development growth of respective countries and the world. Nevertheless, challenges persist in sustainable reporting, such as varying standards, limited resources, required skills, technology for interpreting diverse sustainability metrics and a lack of management support. Furthermore, disclosing environmental efforts might impact an organization's market image, causing some with weak environmental performance to avoid reporting. The ESG framework, with multiple organizations offering guidance and evaluation, lacks a unified structure due to the absence of mandatory standards. Instead, various organizations provide voluntary standards encompassing both quantitative and qualitative disclosures. International guidelines offer overarching frameworks, focusing on areas and governance processes for consideration. Rating agencies use data from self-disclosed and third-party ESG information to calculate ratings, contributing to an organization's score. According to [9], sustainability reporting is a sophisticated information system that reflects the costs associated with the enterprise's and organization's sponsorship and charitable programs, as well as the costs associated regarding environmental and ecological conservation measures that affect the enterprise's and organization's shareholders, clients and regulators. According to [10, 11], in order to allow stakeholders to objectively evaluate the enterprise's or organization's performance during the reporting period, the details provided in the report must be comprehensive to show the impact of their activities on the economy, environment and society. Using information technology to automate accounting operations is the most efficient approach for supplying financial and non-financial data. Information technology is defined as the method utilized to store, process, transmit, secure and display information with the goal of enhancing activity efficiency. It is important to consider the fact that the data is provided by the company's accounting department when setting up an automated sustainability reporting system. As a result, they choose which primary sources will be used to create the reports. According to [9], social activity accounting is a series of activities with the goal of producing a report using data from primary accounting. Automation of accounting operations using information technology is the most efficient method of supplying accounting data [12]. It is crucial that a structured process to govern internal and external communications, including regulatory and mandatory communications (e.g., SEC filings, CRA reports, 10-K forms) is in place to enable automated sustainability reporting system. As a result, they select the primary sources that will be combined to create the reports [13]. AI finds widespread use in sustainability, especially in operational aspects and diversity, equity, and inclusion (DEI) efforts. Additionally, AI assists biodiversity start-ups in safeguarding and restoring forests.

The primary aim of this research is to look into scholarly work done within the academic realm around the role of information technology in sustainable development reporting and understanding the trajectory of development. This investigation also delves into the extent of this trend in middle- and lower-income countries. Given that the subject of "information technology's impact on sustainable development reporting" is still relatively new, this study seeks to explore and analyze key inquiries outlined in *Table 1* of the methodology section. This study has employed the Scopus Database as the primary resource for gathering information. The database was utilized to collect data encompassing research articles, conference proceedings, reviews, book chapters and books related to the keywords "information technology" and "sustainable development reporting." The data collected spanned the years from 1989 to 2023. Additionally, a bibliometric investigation was conducted, encompassing both descriptive and network visualization analyses, using the data extracted from the Scopus database.

1. Literature review

As sustainability is being embedded into corporate strategy, ESG reporting is no longer a voluntary activity; it is becoming a strategic opportunity to transform their value chain to create competitive advantage.

The implications of this decision extend well beyond publishing an annual report or complying with mandatory regulatory requirements, such as the SEC's forthcoming climate disclosure rule or the European Union's Corporate Sustainability Reporting Directive (CSRD). Companies can use this technology to gain insights into their supply chains, to spot methods for reducing a product's carbon footprint or to track progress on net zero targets. Over time, these efforts can potentially have a positive impact on financial performance. One important step toward sustainability reporting in the EU is the Non-Financial Reporting Directive (2014/95/EU) [14] passed in 2014 and applied for the first time in 2017. In 2021, approximately 11,700 large European public-interest companies were required to publish a sustainability report, formally labelled a "non-financial statement" [14]. However, a majority of lower-middle income and lower income countries either do not have a similar kind of regulatory framework or if they have something like this, it is poorly implemented and practiced.

Regulatory bodies are recommending new climate-related disclosures that will steer major transformations and conformity for uplifting the standard of disclosures that registrants make about climate-related risks, their climate-related targets and goals, their greenhouse gas (GHG) emissions and how the board of directors and management handle climate-related risks. The proposal would also require registrants to quantify the effects of certain climate-related events and transition activities in their audited financial statements. They are having major implications for businesses. There has been an increased rigor for process and controls related to GHG measurement and reporting: to include risk assessment that looks into future impacts on strategy and outlook. Enhanced transparency about how the business plans to achieve climate-related goals and details about the specific governance model to drive towards those goals are increasingly being embedded in the corporate vision. The auditing requirement of the financial statements has been expanded to include climate-related impact on Third-party assurance of Scope 1 and Scope 2 emissions required.

Generally, there has been a lack of research on sustainability reporting and this can be attributed to difficulty in parsing unstructured data and the lack of standards for the disclosure. In 2021, the International Sustainability Standards Board (ISSB) was established to create a global baseline of sustainability-related disclosure standards. Along with that, regulatory authorities in countries are also introducing mandatory ESG requirements for certain types of companies. One of the explicit objectives of the ISSB is "to facilitate the addition of requirements that are jurisdiction-specific or aimed at a broader group of stakeholders" [15]. Despite making headway, ESG reporting remains a challenging subject. While mandatory disclosure might seem beneficial from an investor's point of view, there are costs associated with that. The key costs are the preparedness cost to gather the essential data, the proprietary cost of disclosing private information and the costs associated with litigation. On the other hand, voluntary disclosure gives firms the option to be selective with the information that is revealed, often painting an incomplete picture. ESG data providers play an important role for the time being but face the same challenges.

There is a wide range of governance structures and ESG owners vary from one company to another. They can include the general counsel office, chief sustainability officer, investor relations, communications, etc. The data that goes into ESG reports comes from systems that are not SOX controlled, and this data is often disorganized. To meet reporting requirements, the quality of the information will need to be increasingly higher, where the entire value chain is involved to find data, process data, compile data, and control data, even if they are not subject matter experts in ESG.

Given rapid regulatory changes, many companies are digitizing their sustainability data to better understand the current state across ESG metrics, risks and develop a robust reporting framework. It is critical to understand metrics reported today and anticipated future reporting needs (e.g., Scope 1,2,3 emissions, energy usage, water usage, waste, living wages, lost time incident rate, human rights violations, etc.) to define in-scope communication channels (e.g., ESG report, proxy and annual report, publicly disclosed investor presentation, etc.). It helps them understand expected insights to enable management to monitor the metric by validating data sources and collection processes (e.g., external vendor, internal, or a mix) for assessing performance. This approach ensures remediation planning and a roadmap for ESG processes while controlling deficiencies and enhancements based on industry leading practices. Companies that look holistically at their external disclosures and align them to their corporate strategy, data and communications will achieve a competitive advantage. Hence, it is quite clear that due to evolution in the regulatory framework imposed by international agencies and countries, it has been imperative for both corporations and organizations, as well as technology providers to combine, connect and link financial and non-financial view metrics into one holistic reporting. While going through different reports published by the concerned parties, it can be observed that many multinational companies operating in developed markets and technology consulting companies are working on sustainability reporting in a systemic manner as opposed to what is observed in a majority of local companies of middle income, lower middle income and low-income countries.

In the academic domain, research work on the role of information technology in sustainable development reporting is rather descriptive. Less than 300 serious pieces of literature have been indexed in the SCOPUS database from 1989 to 2023. However, since the last decade, research work in this domain has been showing a good increment, as companies all around the world are adopting sustainability reporting obligations as part of their operations to address stakeholder demands. Researchers and authors such as [16–22], underscored the role of corporate sustainable development reporting in companies and published their work concerned with this domain. It is similar in the area of socially and environmentally sustainable development reporting which encompasses the non-financial reporting practice for both corporate and government organizations. A few researchers and academicians such as [23–26], focused on this area and published serious literature. There are many other articles which have been published in the last decade, however, as per the urgency of the curbing climatic degradation for improving the condition of the economy, society and the environment, it is imperative to boost and strengthen research in information technology for developing viable and cost-effective tools for sustainable development monitoring and reporting for companies, academic institutions and government bodies in the low and emerging economies. The present day challenge is to understand how to go from commitment to action and master the complexity of your sustainable transformation. Various business sustainability programs at leading institutions are helping organizations at every step of the way, from "Risk Identification & Assessment," to "Policies, Framework, & Governance," "Risk Reporting," "Data & Technology Strategy and Planning of Execution and Management of Sustainable Reporting". In a variety of company areas, look for prospects for sustainable business models. Discover best practices from renowned, international sustainability firms. Utilize the most up-to-date resources, frameworks, and technologies to place sustainability at the core of your company. With the assistance of instructors, coaches, and peers, create your individual sustainable transformation strategy.

2. Methodology

The Analytical Framework in the Methodology part serves as an analysis manual, making it easier for students to comprehend the entire research process. Each step that must be taken for this investigation is described in this section. This research aims to conduct descriptive and bibliometric analysis in the area of research work on the development and the role of information technology on sustainable development reporting. The research questions are listed in *Table 1* alongside the justifications and analysis methods. In order to give scholars a better understanding of the development in connected subjects, the descriptive analysis gives general information on the annual production, annual citations, and performance of countries, journals, authors and keywords. Apart from descriptive analysis, Bibliometric analysis has also been employed. This is a quantitative method for assessing the bibliographic data in articles and journals. The method is frequently employed to look-into the references to scientific papers that are cited in a journal, to map the journal's scientific field and to categorize research articles by research area. It uses a scientific computer-assisted review process to examine all the publications on a particular subject or area in order to find key authors or pieces of research, as well as their connections [27]. Bibliometric analysis is applicable to various research domains across multiple subject matter by deploying various search approaches and data analysis algorithms. The bibliometric analysis's findings are then provided, including the co-author, co-word and bibliographic coupling analyses, as well as general patterns in publication.

The bibliometric analysis method can alternatively be described as one that follows suggestions [28]. The procedure is used purposefully and adheres to predetermined stages, making it possible for other researchers to replicate it. Bibliographic analysis focuses on quantitative methods to analyse books, journal articles and other written materials. It is frequently used across a wide range of fields [29]. The bibliometric approach includes the application of quantitative techniques on bibliometric data and gathers the bibliometric and intellectual structure of a topic by examining the relationships between several research components [30]. With the use of such data, it is possible to highlight the contributions of a particular field of study, spot links and information silos, as well as trends and prospective gaps [31].

Below is a thorough explanation of each subsection of bibliometric analysis:

1. **Publication analysis.** Estimating the authors' contributions in related disciplines using the complete counting methodology, which correlates full recognition for related contributions [32].
2. **Citation analysis.** Determines an article's popularity by counting how often it is mentioned [33].
3. **Co-authorship analysis.** Analyzes the frequency of joint publications to track national cooperation efforts [34].
4. **Co-word / co-occurrence analysis.** Discover research hotspots via the degrees of keywords co-occurrence [35].
5. **Bibliographic coupling.** Find the bibliographic connections between two publications [36].

In this study, a procedure was established for choosing the search words, database to use, selection criteria for the search, software to use for analysis and results analysis. *Figure 1* below illustrates these steps, and the following paragraphs go into greater depth about them.

Definition of search terms. The purpose of the current article, as stated in the introduction, is to (I) identify and analyze the nature and evolution of literature related to information technology and cohesion with sustainable development reporting, and (II) provide a comprehensive systematic review on the development of the research work related to information technology and cohesion with sustainable development reporting and their connection to specific disciplines.

Selection of database and data collection strategy. The subsequent keywords were searched under the criteria of Title, Keywords and Abstract of the publications by using the search strings connected to information technology and sustainable development reporting: TITLE-ABS-KEY ("information technology") and TITLE-ABS-KEY ("sustainable development reporting") to collect articles' data in an electronic database.

Table 1.

Overview of the research development in this study

Descriptive / Bibliometric analysis			Research methodology
No.	**Research question(s)**	**Research objective(s) and aims**	
1	What is the publication trend of literature on the role of IT on SDR?	To understand how the development and role of IT in the area of SDR study has evolved over the years. This objective is essential to assist researchers in visualizing the potential of IT in the evaluation of SDR.	Descriptive analysis (publication analysis; citation analysis).
2	Which countries contributed to the development and the role of IT on SDR?	To identify the countries that contributed the most and received the most citations for their work. To promote international research collaboration, it is essential to achieve this goal.	Descriptive analysis (publication analysis; citation analysis).
3	Which journals led in the field of the development and the role of IT on SDR research?	To identify the journals that published the most articles on the relationship between IT and SDR. This goal is crucial in assisting researchers in selecting platforms for releasing and communicating their findings.	Descriptive analysis (publication analysis; citation analysis).
4	Which are the influential authors on the development and the role of IT on SDR research?	Find out which papers in the linked field are the most read. This goal is important since it might help researchers identify the research gaps in the associated papers.	Descriptive analysis (citation analysis).
5	How is the countries' collaboration structure in the area of the development and the role of IT on SDR research?	To assess the countries' collaboration trend in the development and role of IT in the area of SDR. This objective aims to help researchers decide which country is suitable for collaborating in publishing in this research area.	Co-authorship analysis.
6	What is the conceptual structure of keywords in the area of the development and the role of IT on SDR research?	To identify the research hotspots that evolved in the field. This objective helps researchers to understand the new research topics.	Conceptual structure analysis (co-occurrence analysis of words).
7	What is the countries' coupling structure in the area of the development and the role of IT on SDR research?	Presenting information on commonalities between two countries. This goal helps to assess the degree to which these countries' ideas and literary works are alike.	Intellectual structure analysis (bibliographic coupling).
8	Which are the research fronts of the development and the role of IT on SDR research study?	To identify papers in the field that have a similar theme. This goal gives scholars a sense of the topics covered in the papers, which helps them when they are creating new research projects.	Intellectual structure analysis (bibliographic coupling).

The search process was conducted based on the abovementioned keywords. Since the Scopus database is one of the largest and most widely accepted scientific databases and has an extensive number of literary materials including peer reviewed articles, conference proceedings, reviews, book chapters, books, etc., it became the primary choice for the current course of investigation. Searching for "information technology" and "sustainable development reporting" in the Scopus database showed up total of 259 documents. Since only popular types of documents were included in this study such as articles from journals, review articles, conferences proceedings, book chapters, and conference reviews, subsequent filtering of the search results was needed. After filtering out short surveys (one document), Note (one document), and Erratum (one document), only 256 documents were retrieved. The first document related to the chosen search string appeared in 1989 and was included in the Scopus database. Hence, the time frame criterion from year 1989 to 2023 was selected for the data collection. While using the Scopus database, the aforementioned measures were followed, and bibliographic data was exported in CSV format without further data cleansing.

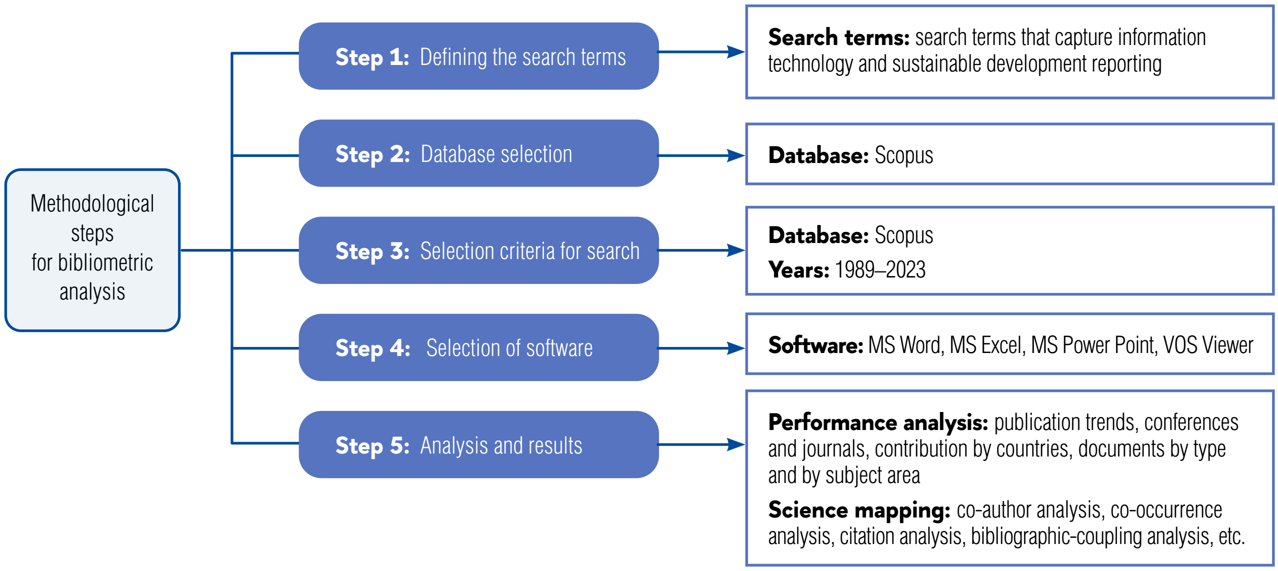


Fig. 1. Methodological scheme for the bibliometric analysis.

Software selection and data analysis. This article's preparation involved the use of numerous Microsoft Office programs. Nevertheless, the management and analysis of the collected data was primarily supported by the using two software programs, namely Microsoft Excel 2020, a popular spreadsheet, and VOS Viewer, a freeware data-visualization program. The search information was examined and categorized based on the annual number of publications, publication journals, contributing authors and productive nations. The next step was visualizing the data on "information technology" and "sustainable development reporting" in terms of documents by year, documents by country, documents per year by source, documents by authors, co-authorship–related countries, co-occurrence–related author keywords, and bibliographic coupling in terms of country and publication.

3. Results

The bibliometric analysis utilizes a tremendous amount of studies to identify popular trends in the development and role of information technology in the domain of sustainable development reporting. Therefore, this section emphasizes the results generated via bibliometric analytic tools.

3.1. Descriptive analysis

In order to analyze current trends in the research of the development and role of information technology on sustainable development reporting, this subsection offers a thorough overview of publication trends and citation performance on this study, followed by the most productive nations and the most important journals and articles. The number of papers on "information technology" and "sustainable development reporting" that were published in Scopus was the first finding based on data extracted from the Scopus database. The total number of documents retrieved was 256.

3.1.1. Documents per year

The following figure (*Fig. 2*) shows the number of documents published annually. The first document related with "information technology" and "sustainable development reporting" was published in 1989. From year 1990 to 2000, the average number of articles published annually was less than five. However, from 2001 to 2010, there was a growth in published research work and this can be observed with one deflection that occurred in 2010. From 2011 onwards, the number of published documents has

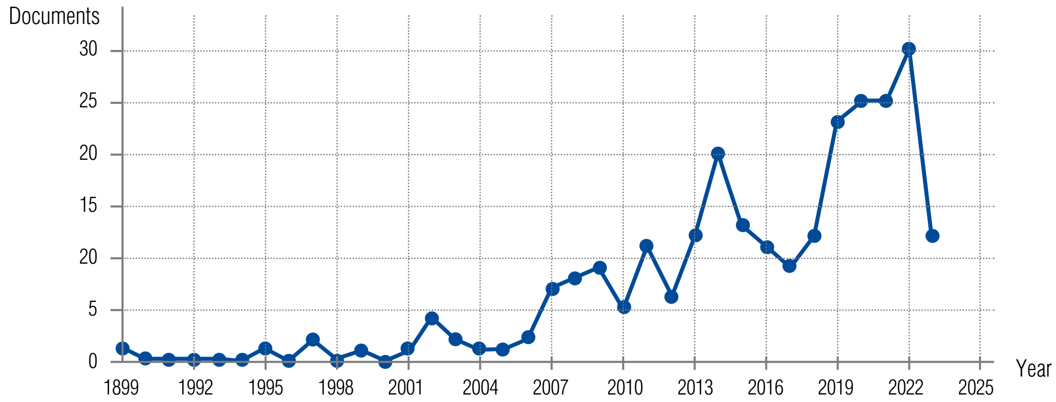


Fig. 2. Documents published annually.

increased by three times as compare to 2010. However, some downturn can be seen in year 2014, 2015 and 2016.

3.1.2. Document by country

As per the following *Table 2*, there are variations in research output and impact among different countries, with factors such as the number of publications, total citations, and average citation indicating the prominence of their research in the given field. It is quite clear that developed countries such as the USA, UK, Canada, Australia lead in both the number of publications and total citations, indicating a strong research output and impact in the field. The high average citation rate suggests that their research is frequently cited by others. On the other hand, despite a lower number of publications, China, amongst the developing countries, has a high total citation count and an exceptionally high average citation, suggesting a significant impact of their research.

3.1.3. Documents per year by source

Table 3 and *Fig. 3* shows the top ten peer-reviewed journals with the most publications in the field of information technology and sustainable development reporting. This analysis indicates variations in publication and citation statistics, average citation and scientific journal rankings among these sources. The SJR values give an idea of how these sources are ranked in terms of their scientific impact within their respective fields. The journal with the most such publication numbers is Sustainability from Switzerland. This source has a moderate number of publications and citations, along with a reasonable average citation. The SJR indicates a moderate journal ranking. This journal Sustainability was published by MDPI AG, then followed by Environmental Science and Engineering Subseries Environmental science and Journal of cleaner production, then followed by Technological forecasting and social change.

Table 2.

The top 10 most productive countries

No.	Country	TP	TC	AC
1	United States of America	39	2461	63.1
2	United Kingdom	31	1889	60.9
3	Canada	20	1574	78.7
4	Australia	20	572	28.6
5	Germany	17	429	25.2
6	India	17	165	9.7
7	Russia	14	50	3.5
8	Italy	13	193	14.8
9	China	12	1696	141.3
10	Spain	12	96	8

Note:
TP indicates the complete publication of articles according to countries; TC is the total citation, while AC is the ratio of total citation per total publication.

Table 3.

The top 10 most influential journalsbased on number of publications

No.	Source	TP	TC	AC	SJR
1	Sustainability (Switzerland) MDPI AG	12	190	15.8	0.664
2	Journal of Cleaner Production	5	355	71	1.981
3	Environmental Science and Engineering (Subseries: Environmental Science)	5	34	6.8	0.125
4	Lecture Notes in Networks and Systems	3	0	0	0.151
5	Technological Forecasting and Social Change	2	18	9	2.644
6	Science of the Total Environment	2	39	19.5	1.946
7	Frontiers In Public Health	2	23	11.5	1.125
8	Sage Open	2	41	20.5	0.462
9	Procedia Engineering	2	22	11	0.185
10	Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis	2	59	29.5	0.169

Note:
TP indicates the complete publication of articles according to countries, TC is the total citation, while AC is the ratio of total citation per total publication, SJR is Scientific Journal Rankings

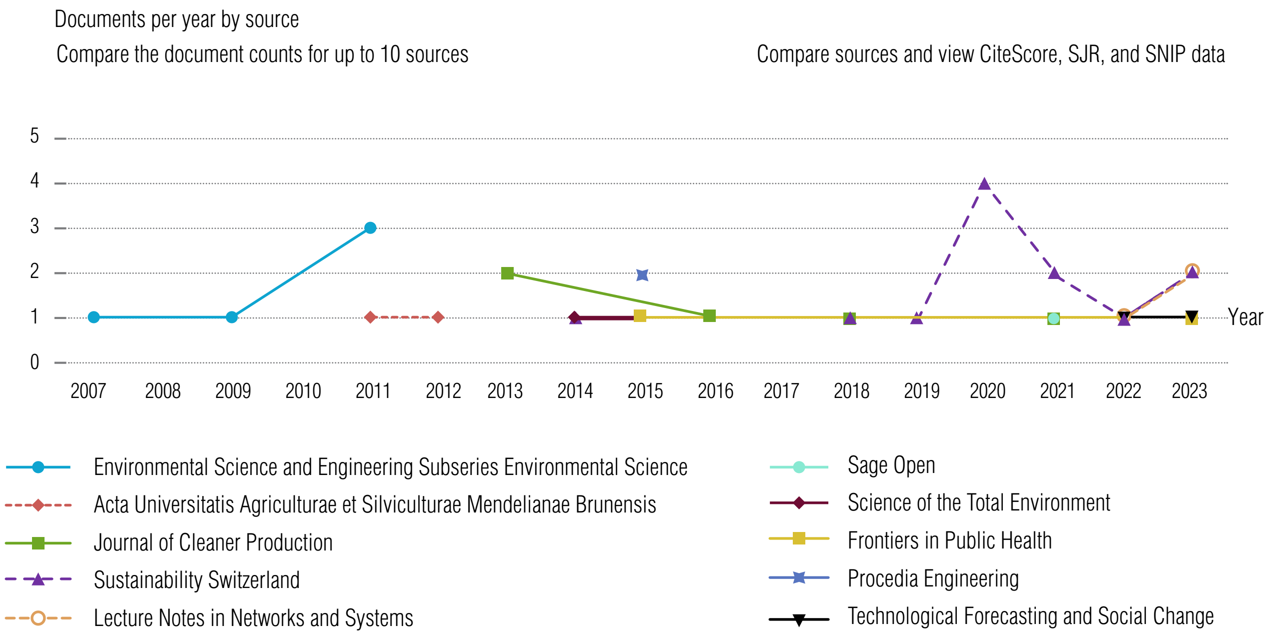


Fig. 3. The top 10 most influential journals based on number of publications.

3.1.4. Documents by authors

Table 4, indicates the top ten contributing authors with published articles in peer-reviewed journals in information technology and sustainable development reporting fields. The top four authors produced three articles each, whereas the bottom six authors produced two articles on this field individually. However, it can be seen that these authors have worked together on the same articles; therefore, the number of articles or proceedings are not increasing significantly.

3.2. Network visualization

Network visualization reflects the co-authorship of countries, co-words, and bibliographic coupling. Network analysis provides researchers with a better graphical visualization on collaboration, co-occurrences, and bibliographic coupling, where the relations between selected items is illustrated using nodes size, nodes color, and the thicknesses of connecting lines [37].

Table 4.

The top 10 most influential authors

No.	Author	Article Name / Conference Proceedings; Year; Total Citation
1	Gómez, J.M.	"Sustainable online reporting model – A web-based sustainability reporting software"; 2011; 1 "Concept and implementation of a flexible and differentiated shopping cart functionality for creating personalized sustainability reports"; 2008; 3 "Conception of system supported generation of sustainability reports in a large-scale enterprise"; 2007;1
2	Süpke, D.	"ProPlaNET – Collaborative sustainable project planning – A comparison with existing approaches"; 2011; 0 "Sustainable online reporting model – A web-based sustainability reporting software"; 2011; 1 "Concept and implementation of a flexible and differentiated shopping cart functionality for creating personalized sustainability reports"; 2008; 3
3	Trenz, O.	"Corporate performance indicators for agriculture and food processing sector"; 2012; 20 "Corporate performance evaluation and reporting"; 2011; 3 "Integration of economic, environmental, social and corporate governance performance and reporting in enterprises"; 2011; 39
4	Štencl, M.	"Corporate performance indicators for agriculture and food processing sector"; 2012; 20 "Corporate performance evaluation and reporting"; 2011; 3 "Integration of economic, environmental, social and corporate governance performance and reporting in enterprises"; 2011; 39
5	Arsenault, A.	"Integrated modelling software platform development for effective use of ecosystem models"; 2015; 3 "Integrated modelling software platform development for effective use of ecosystem models"; 2014; 8
6	Bhatti, J.	"Integrated modelling software platform development for effective use of ecosystem models"; 2015; 3 "Integrated modelling software platform development for effective use of ecosystem models"; 2014; 8
7	Braun, P.	"Intelligent mortality reporting with FHIR"; 2018; 8 "Intelligent mortality reporting with FHIR"; 2017; 5
8	Chokkavarapu, N.	"Role of drone technology in sustainable rural development: Opportunities and challenges"; 2023; 0 "Role of drone technology in sustainable rural development: Opportunities and challenges"; 2022; 0
9	Glebkova, I.Y.	"Technology application of the concept of market-oriented reporting in accounting and statistics"; 2019; 0 "Technology for determining the effectiveness of representative offices of the companies abroad"; 2019; 1
10	Hartcher, M.G.	"A governance framework for data audit trail creation in large multi-disciplinary projects"; 2013; 1 "Driving data management cultural change via automated provenance management systems"; 2013; 1

3.2.1. The co-authorship network of documents

Table 5 and *Fig. 4* display the co-authorship map in the "information technology" and "sustainable development reporting" articles. There were 91 countries according to VOS viewer. However, when a filter was applied to sources having at least one document per country with a minimum of one citation, then 72 documents met the threshold. The USA, along with the UK, Canada and Australia have a high publication count and a substantial number of links, indicating strong research collaboration with a significant link strength. On the other end, Russia has a lower publication count and zero links and link strength, with no collaboration in this dataset. The other counties in the scope, showed a moderate publication count and link strength, suggesting active research participations. Overall, the analysis indicates variations in publication count, research collaboration through links, and link strength among different countries. Countries with higher publication counts and stronger link strengths are likely more active in research collaboration within the dataset. The absence of links and link strength for Russia suggests limited representation in this specific dataset's collaborative network. The network visualization result also shows a number of nodes and clusters. The distance between the nodes and the link thickness reveals the degree of cooperation between countries, while the links between the nodes represent the relationships between the countries, for example, a red cluster showing strong connection among Canada, China, Indonesia, etc. Similarly, a green cluster shows strong connection among Greece, Turkey, Romania, etc.

3.2.2. The co-occurrence network of keywords in documents

Keywords co-occurrence analysis is a popular platform used in bibliometric analysis, since it helps reveal the core research topic. To do so, a VOS viewer was employed and the analysis result displayed the keyword distribution map in the "information technology" and "sustainable development reporting" articles based on the author keywords. Once 852 keywords were filtered using a minimum appearance criterion of three times, the output returned 35 keywords. Keywords that are

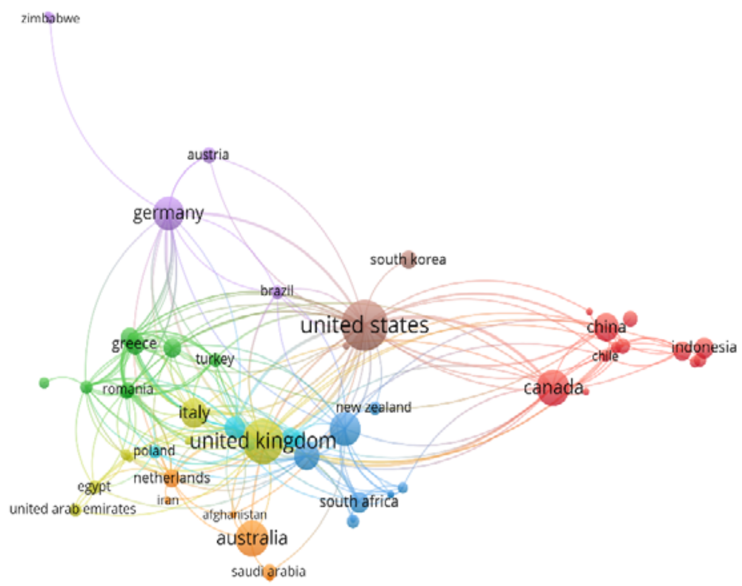


Fig. 4. The co-authorship network of articles on "information technology" and "sustainable development reporting" in terms of countries.

Table 5.

The top 10 most influential countries in terms of co-authorship network of articles on "information technology" and "sustainable development reporting"

No.	Country	TP	Link	TLS
1	United States of America	39	34	52
2	United Kingdom	31	35	55
3	Canada	20	16	21
4	Australia	20	6	7
5	Germany	17	20	25
6	India	17	19	24
7	Russia	14	0	0
8	Italy	13	23	33
9	China	12	11	13
10	Spain	12	22	28

Note:
TP indicates the complete publication of articles according to countries, TLS indicates Total Link Strength

Table 6.

The co-occurrence network of keywords in articles on "information technology" and "sustainable development reporting"

No.	Keywords	Occurrence	Link	TLS
1	Sustainable development	18	11	19
2	Sustainability	17	10	11
3	Sustainability reporting	14	10	24
4	Information technology	11	5	6
5	Reporting	9	10	14
6	GIS	5	1	1
7	GRI	4	7	19
8	Digitalization	4	5	7
9	Information systems	4	5	5
10	Climate change	4	3	3

Note:
TLS indicates Total Link Strength

unconnected to each other were not included in the analysis. *Table 6* highlights the top 10 authors' keywords arranged according to the occurrence and total link strength. "Sustainable development" recorded the highest occurrence (18) and total link strength (19), followed by "sustainability" (17) and (11). After that, "sustainability reporting" appeared (14) times with total link strength of (24), followed by "information technology" (11) and (6), and finally keyword "reporting" with (9) appearances and having (14) total link strength. These keywords also ranked as the top five most popular authors' keywords as seen in the descriptive analysis. Overall, the analysis reveals variations in the occurrence of keywords, their linkages and link strength. Some terms have stronger linkages and connections, while others occurred on account of their stronger mutual associations within the dataset. The significance of these observations depends on the context in which the keywords are being analyzed.

Concurrently, the co-occurrence of authors' keywords can be visualized via network mapping, as shown in *Fig. 5*. The size and color of nodes play different roles in the co-occurrence analysis, where the sizes reflect the frequency of the authors' keywords on the role of information technology in sustainable development reporting. It can be seen in *Fig. 5*, that keyword "sustainable development"; belongs to biggest node followed by "sustainability," "sustainability reporting," "information technology," "reporting," etc. A link connecting two terms shows that they appeared together, and the thickness of the link shows how frequently they occurred together. Additionally, the distance between nodes indicates the degree of their interaction between keywords.

3.2.3. Bibliographic coupling

According to [38], bibliographic coupling was initially used by [39] to identify the connections between two articles. Bibliographic coupling plays a prominent role in determining the relatedness of selected items such as

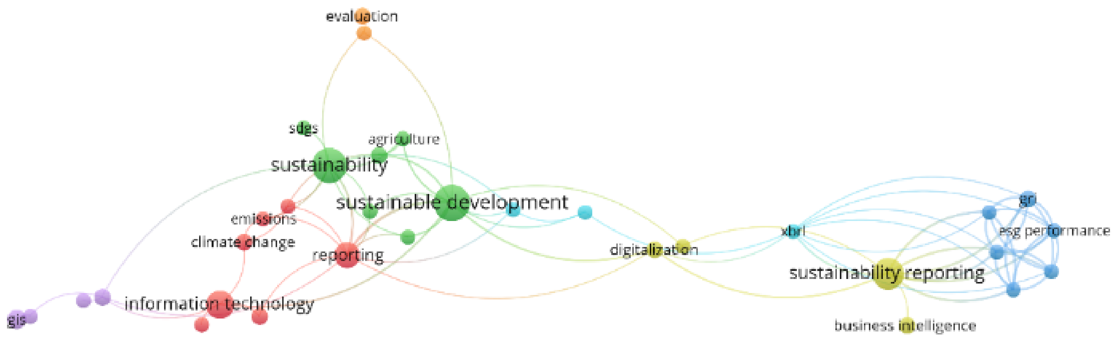


Fig. 5. The co-occurrence network of keywords in articles on “information technology” and “sustainable development reporting”.

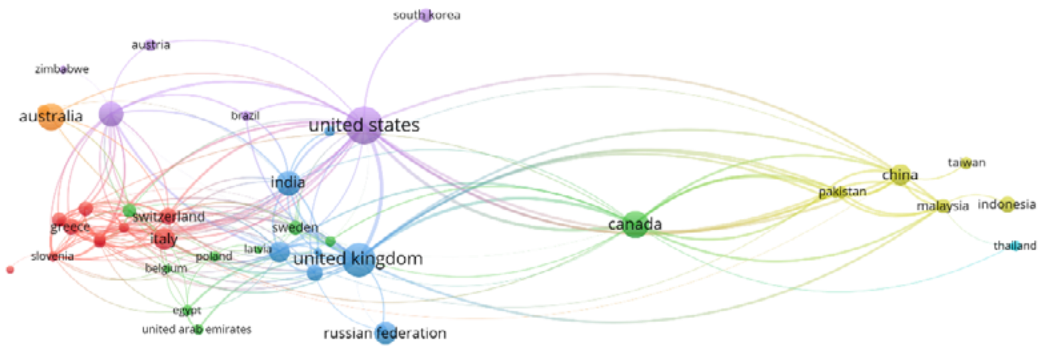


Fig. 6. The bibliographic coupling network of countries on “information technology” and “sustainable development reporting”.

countries and publications. In this work, the bibliographic coupling of geographic regions and publications was carried out.

3.2.3.1. Bibliographic coupling of countries

Bibliographic coupling of countries occurs when an article from two countries cites another document in its reference list [40]. Bibliographic coupling illustrates the frequency with which these countries have common references in their bibliographies, enabling us to gauge the resemblance between these publications. The graphical representation of bibliographic coupling can be displayed using different combinations of colors and sizes of nodes. The colors signify the quantity of clusters identified in this analysis, while the size of nodes depict the contributions of each country – larger nodes denote more noteworthy contributions from the respective country. The investigation of bibliographic coupling was the most recent meta-analysis. 47 of the 91 countries met the requirement of having at least two documents and one citation per country.

Similar to the results in *Table 2*, the coupling analysis showed that the United States of America, United Kingdom, Canada, Australia and Germany are the top five countries in studies involving the research on the development and role of information technology on sustainable development reporting. The bibliographic coupling of countries involved seven clusters with distinct colors as seen in *Fig. 6*. The Italy leads the densest cluster (red cluster) and is closely coupled with Switzerland, France, Greece, Slovenia, etc. United Kingdom leads the second-dense cluster (blue cluster) and is strongly coupled with India, Spain, Russia and South Africa. The United States of America (purple cluster), China (yellow cluster) and Canada (green cluster) also act as the leading countries in terms of cluster size. However, if in terms of overall analysis of coupling networks, it can be inferred that United States of America, United Kingdom, and Canada play the vital role in the research on the development and role of information technology in the sustainable development reporting, as many countries are coupled with these countries. However, despite a lower number of publications, China, too, is seen emerging as an active country in this area of research.

3.2.3.2. Articles

Bibliographic coupling occurs when two publications share a common reference article. This implies that articles with similar research focuses can be identified by examining the bibliographic coupling of their respective publications. The exploration of bibliographic coupling was carried out in the most recent meta-analysis. Out of 256 documents, 153 fulfilled the criterion of having at least two citations per document. In this analysis, the relationship between the information technology and developments in areas of sustainable development reporting is represented by colors and node sizes. The node colors indicate that clusters existed on this topic, while node sizes represent the total citations gained by a paper. As per the *Fig. 7*, the role of information technology in sustainable development reporting publications were grouped into two clusters by red and green node colors. The red cluster comprises five publications and a total of 76 citations, with an average of 15 citations per article whereas the green cluster has 4 publications and a total of 72 citations, with an average of 18 citations per article. The most cited publication in this (red) cluster focuses on "Internet-supported sustainability reporting: Developments in Germany" [40] (51 citations), followed by "Corporate sustainability reporting and disclosure on the web: An exploratory study" [20] (7 citations). The most cited publication in the green cluster focuses on the "Integration of economic, environmental, social and corporate governance performance and reporting in enterprises" [19] (39 citations), followed by "Corporate performance indicators for agriculture and food processing sector" [41] (20 citations).

Although the publications covered diverse topics, these clusters generally have comparable research areas. Therefore, it can be said that the "Development and role of information technology in sustainable development reporting" publications are ideal for use in sustainable socio-economic and environmental development.

4. Discussion

The above findings offer a detailed analysis of bibliographic traits in globally recognized Scopus-indexed journals. These articles center around "information technology" and "sustainable development reporting," alongside related fields such as "machine learning" and "sustainable development reporting" and "big data" and "sustainable development reporting." These areas have attained significant and at times unprecedented traction, especially in the private sector, though academic explora-

Fig. 7. The bibliographic coupling network of publications on "information technology" and "sustainable development reporting".

tion is rather limited, which is evident by scarce articles and conference presentations. The journals and conferences selected underscore the global scientific community's limited interest in these subjects and this trend can be seen in the descriptive and network visualization results obtained in the current course of research work. Co-authorship relationships analysis scope involved researchers from 90+ countries, with intense collaborations. However, the major collaborators belong to developed countries such as the United States of America, the United Kingdom, Australia, Canada, Germany, etc. Keyword co-occurrence patterns reveal associations, pivotal in guiding new research trends. Recent research topics are shown to emphasize "sustainable development," followed by "sustainability," "information technology" and "reporting" for future work. This signifies a close relationship between role of IT in sustainable development and the growing importance of transparent ESG reporting addressing interests of investors and regulators. Despite the number of publications, there is no direct correlation between them and citations. Notably, Sustainability (Switzerland), The Journal of Cleaner Production, and Environmental Science hold citation prominence. A few authors have been found to show high co-citation frequency, indicating widespread sources of information. However, limitations exist – this study focuses on specific units and high-level ESG terms, which might have potentially excluded other relevant articles. Future studies could diversify units, incorporate specific keywords and expand searches across broader databases for a wider research scope.

Conclusion

The study's statistical findings reveal a notable increase in the number of articles focused on sustainability reporting published in Scopus-indexed journals, especially in the last decade. The research emphasizes growing opportunities for collaboration among authors from different countries. Additionally, the study points out the abundance of easily accessible sources related to Environmental, Social, and Governance (ESG) reporting, indicating a potential for future research. As climate change impact become more tangible globally, all topics impacting sustainability will become crucial for both corporations and society. The analysis indicates that research on the role of information technology in sustainable development reporting is concentrated in countries with higher income levels, while middle and lower-middle-income countries have contributed less. The study underscores the need for comprehensive ESG reporting that creates a coherent narrative for various stakeholders. Future research frameworks should focus on sustainability goals, standardizing reporting content according to guidelines, and exploring links between ESG practices and United Nations Sustainable Development Goals (SDGs). To achieve a sustainable future, active IT consulting firms, companies, organizations, and governments must provide expertise and affordable solutions for sustainable development reporting. Encouraging middle and lower-middle-income countries to adopt such approaches can contribute to sustainable development. Increased research in this field can enhance the understanding and impact of sustainable business practices, influencing positive environmental, social, and governance outcomes, and creating long-term stakeholder value. This assistance will facilitate the participation of these countries, encouraging them to invest in technological infrastructure and non-financial accounting frameworks for effective and standardized sustainable development reporting practices. ■

References

1. Brundtland G. (1987) *Report of the World Commission on Environment and Development: Our common future*. United Nations General Assembly document A/42/427.
2. Boons F., Lüdeke-Freund F. (2013) Business models for sustainable innovation: State-of-the-art and steps towards a research agenda. *Journal of Cleaner Production*, vol. 45, pp. 9–19.
3. Bocken N., Short S., Rana P., Evans S. (2014) A literature and practice review to develop sustainable business model archetypes. *Journal of Cleaner Production*, vol. 65, pp. 42–56.

4. Sachs J.D. (2012) From millennium development goals to sustainable development goals. *Lancet*, vol. 379, pp. 2206–2211.
5. Agarwal N., Gneiting U., Mhlanga R. (2017) *Raising the bar: Rethinking the role of business in the sustainable development goals*. Oxford: Oxfam, 2017.
6. Busco C., Frigo M.L., Quattrone P., Riccaboni A. (2013) *Towards integrated reporting: Concepts, elements and principles.* London: Springer, 2013.
7. Dumay J., Bernardi C., Guthrie J., DeMartini P. (2016) Integrated reporting: A structured literature review. *Accounting Forum*, vol. 40(3), pp. 166–185. https://doi.org/10.1016/j.accfor.2016.06.001
8. Simnett R., Huggins A. (2015) Integrated reporting and assurance: Where can research add value? *Sustainability Accounting, Management and Policy Journal*, vol. 6(1), pp. 29–53. https://doi.org/10.1108/SAMPJ-09-2014-0053
9. Fakhretdinova E.N., Klychova G.S., Klychova A.S., Antonova N.V. (2015) Development of accounting and financial reporting for small and medium-sized businesses in accordance with international financial reporting standards. *Asian Social Science*, vol. 11(11), pp. 318–322.
10. González M., del Mar Alonso-Almeida M., Avila C., Dominguez D. (2015) Modeling sustainability report scoring sequences using an attractor network. *Neurocomputing*, vol. 168, pp. 1181–1187.
11. Alonso-Almeida M., Llach J., Marimon F. (2014) A closer look at the 'Global Reporting Initiative' sustainability reporting as a tool to implement environmental and social policies: A worldwide sector analysis. *Corporate Social Responsibility and Environmental Management*, vol. 21(6), pp. 318–335.
12. Klychova G.S., Ziganshin B.G., Zakirova A.R., Valieva G.R., Klychova A.S. (2017) Benchmarking as an efficient tool of social audit development. *Journal of Engineering and Applied Sciences*, vol. 12, pp. 4958–4965.
13. Klychova G.S., Zakirova A.R., Zakirov Z.R., Iskhakov A.T. (2014) Development of primary accounting of crop farming products arrival. *Bulletin of Kazan State Agrarian University*, vol. 34(4), pp. 23–28.
14. IFRS Foundation (2021) *IFRS Foundation announces International Sustainability Standards Board, consolidation with CDSB and VRF, and publication of prototype disclosure requirements*. Available at: https://www.ifrs.org/news-and-events/news/2021/11/ifrs-foundation-announces-issb-consolidation-with-cdsb-vrf-publication-of-prototypes/ (accessed 1 November 2023).
15. European Commission (2021) *Proposal for a Directive of the European Parliament and of the Council amending Directive 2013/34/EU, Directive 2004/109/EC, Directive 2006/43/EC and Regulation (EU) No 537/2014, as regards corporate sustainability reporting*. COM/2021/189 final, 2021/0104(COD), Brussels, 21.4.2021. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0189 (accessed 1 November 2023).
16. Chvatalová Z., Kocmanová A., Dočekalová M. (2011) Corporate sustainability reporting and measuring corporate performance. *Environmental Software Systems. Frameworks of eEnvironment. ISESS 2011. IFIP Advances in Information and Communication Technology* (eds. J. Hřebíček, G. Schimak, R. Denzer), vol. 359, pp. 245–254. https://doi.org/10.1007/978-3-642-22285-6_27
17. Medel F., García L., Enriquez S., Anido M. (2011) Reporting models for corporate sustainability in SMEs. *Information Technologies in Environmental Engineering. Environmental Science and Engineering* (eds. P. Golinska, M. Fertsch, J. Marx-Gómez), vol. 3. https://doi.org/10.1007/978-3-642-19536-5_32
18. Hřebíček J., Štencl M., Trenz O., Soukopová J. (2011) Corporate performance evaluation and reporting. Proceedings of the *11th WSEAS international conference on Applied informatics and communications,* and Proceedings of the 4th *WSEAS International conference on Biomedical electronics and biomedical informatics,* and Proceedings of the *International conference on Computational engineering in systems applications*, pp. 338–343.
19. Hřebíček J., Soukopová J., Štencl M., Trenz O. (2011) Integration of economic, environmental, social and corporate governance performance and reporting in enterprises. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, vol. 59(7), pp. 157–166.

20. Raghupathi V., Raghupathi W. (2019) Corporate sustainability reporting and disclosure on the web: An exploratory study. *Information Resources Management Journal*, vol. 32(1), pp. 1–27.

21. Skouloudis A., Malesios C., Dimitrakopoulos P.G. (2019) Corporate biodiversity accounting and reporting in mega-diverse countries: An examination of indicators disclosed in sustainability reports. *Ecological Indicators*, vol. 98, pp. 888–901. https://doi.org/10.1016/j.ecolind.2018.11.060

22. Bruant R., Bennett K., Fox S., Willard S., Michel A. (2021) ESG reporting in the oil and gas industry – A permian basin water management perspective. Proceedings of the *9th Unconventional Resources Technology Conference*. URTeC: 5325. https://doi.org/10.15530/urtec-2021-5325

23. Erin O.A., Bamigboye O.A., Oyewo B. (2022) Sustainable development goals (SDG) reporting: An analysis of disclosure. J*ournal of Accounting in Emerging Economies*, vol. 12(5), pp. 761–789. https://doi.org/10.1108/JAEE-02-2020-0037

24. Mulyani S., Kasim E., Sudrajat (2017) Financial reporting quality and public accountability in regional government: Analysis for the impact of competence, internal control and information technology, Indonesia evidence. Proceedings of the *30th International Business Information Management Association Conference, IBIMA 2017, Madrid, Spain*.

25. Ibatova A.Z., Sitdikov F.F., Klychova G.S. (2018) Reporting in the area of sustainable development with information technology application. *Management Science Letters*, vol. 8(7), pp. 785–794.

26. van Dijk A., Mount R., Gibbons P., Vardon M., Canadell P. (2014) Environmental reporting and accounting in Australia: Progress, prospects and research priorities. *Science of The Total Environment*, vols. 473–474, pp. 338–349. https://doi.org/10.1016/j.scitotenv.2013.12.053

27. de Bellis N. (2009) *Bibliometrics and citation analysis: From the science citation index to cybermetrics*. Scarecrow Press.

28. Block J.H., Fisch C. (2020) Eight tips and questions for your bibliographic study in business and management research. *Management Review Quarterly*, vol. 70, pp. 307–312.

29. Garza-Reyes J.A. (2015) Lean and green-a systematic review of the state-of the art literature. *Journal of Cleaner Production*, vol. 102, pp. 18–29. https://doi.org/10.1016/j.jclepro.2015.04.064

30. Heersmink R., van den Hoven J., van Eck N. J., van den Berg J. (2011) Bibliometric mapping of computer and information ethics. *Ethics and Information Technology*, vol. 13(3), pp. 241–249. https://doi.org/10.1007/s10676-011-9273-7

31. Donthu N., Kumar S., Mukherjee D., Pandey N., Lim W.M. (2021) How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, vol. 133, pp. 285–296.

32. Waltman L., van Eck N.J. (2015) Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, vol. 9, pp. 872–894.

33. Ding Y., Cronin B. (2011) Popular and/or prestigious? Measures of scholarly esteem. *Information Processing & Management*, vol. 47, pp. 80–96.

34. Caviggioli F., Ughetto E. (2019) A bibliometric analysis of the research dealing with the impact of additive manufacturing on industry, business and society. *International Journal of Production Economics*, vol. 208, pp. 254–268.

35. Bui T.D., Ali M.H., Tsai F.M., Iranmanesh M., Tseng M.L., Lim M.K. (2020) Challenges and trends in sustainable corporate finance: A bibliometric systematic review. *Journal of Risk and Financial Management*, vol. 13, 264.

36. Abdullah S., Naved Khan M. (2021) Determining mobile payment adoption: A systematic literature search and bibliometric analysis. *Cogent Business & Management*, vol. 8, 1893245.

37. Zhang G., Kang L., Gu D., Wang X., Yang X., Zhu K., Liang G. (2019) Visualizing knowledge evolution and hotspots of rural environment and health: A systematic review and research direction. *IEEE Access*, vol. 7, pp. 72538–72550.

38. Freire R.R., Veríssimo J.M.C. (2020) Mapping co-creation and co-destruction in tourism: A bibliographic coupling analysis. *Anatolia*, vol. 32(2), pp. 207–217.

39. Kessler M.M. (1963) Bibliographic coupling between scientific papers. *American Documentation*, vol. 14, pp. 10–25.

40. Gu Z., Meng F., Farrukh M. (2021) Mapping the research on knowledge transfer: A Scientometrics approach. *IEEE Access*, vol. 9, pp. 34647–34659.

41. Herzig C., Godemann J. (2010) Internet-supported sustainability reporting: developments in Germany. *Management Research Review*, vol. 33(11), pp. 1064–1082. https://doi.org/10.1108/01409171011085903

42. Hřebíček J., Popelka O., Štencl M., Trenz O. (2012) Corporate performance indicators for agriculture and food processing sector. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, vol. 60(4), pp. 121–132. https://doi.org/10.11118/actaun201260040121

About the authors

Maneesh Kumar Pandey

Doctoral Student, Faculty of Technological Management and Innovations, ITMO University, 9, Lomonosov str., St. Petersburg 191002, Russia;

E-mail: maneeshban@gmail.com

ORCID: 0000-0002-7789-4247

Amit Kumar Pathak

Ph.D.;

Assistant Professor, University of Technology and Applied Sciences – Muscat, PO Box 74, Al-Khuwair, Muscat 133, Sultanate of Oman;

E-mail: Amit.Pathak@utas.edu.com

ORCID: 0000-0003-4777-5180

Irina G. Sergeeva

Dr. Sci. (Econ.);

Professor, Faculty of Technological Management and Innovations, ITMO University, 9, Lomonosov str., St. Petersburg 191002, Russia;

E-mail: igsergeeva@gmail.com

ORCID: 0000-0001-7314-7765